



**UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE BELÉM
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
MESTRADO ACADÊMICO EM ENGENHARIA ELÉTRICA**

CEZAR AUGUSTO FIGUEIREDO DA ROCHA

**TÍTULO: PREVISÃO A CURTO PRAZO DE PREÇO DE ENERGIA UTILIZANDO
MODELO HÍBRIDO COM SARIMAX E XGBOOST**

NÚMERO DA DEFESA: DM 10/2025

**BELÉM
2025**

CEZAR AUGUSTO FIGUEIREDO DA ROCHA

**TÍTULO: PREVISÃO A CURTO PRAZO DE PREÇO DE ENERGIA UTILIZANDO
MODELO HÍBRIDO COM SARIMAX E XGBOOST**

NÚMERO DA DEFESA: DM 10/2025

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica - PPGEE, do Campus de Belém, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Área de concentração: Sistemas de Elétricos de Potência

Orientador(a): Dr(a). Maria Emilia de Lima Tostes

Coorientador(a): Dr(A). Suzane Cruz de Aquino Monteiro

**BELÉM
2025**

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

D111p Da Rocha, Cezar Augusto Figueiredo.
PREVISÃO A CURTO PRAZO DE PREÇO DE ENERGIA
UTILIZANDO MODELO HÍBRIDO COM SARIMAX E
XGBOOST / Cezar Augusto Figueiredo Da Rocha, . — 2025.
61 f. : il. color.

Orientador(a): Prof^a. Dra. Maria Emilia de Lima Tostes
Coorientação: Prof^a. Dra. Suzane Cruz de Aquino Monteiro
Dissertação (Mestrado) - Universidade Federal do Pará,
Instituto de Tecnologia, Programa de Pós-Graduação em
Engenharia Elétrica, Belém, 2025.

1. Previsão de preços de energia. 2. Modelos híbridos. 3.
SARIMAX-XGBoost. I. Título.

CDD 621.3191



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA


“PREVISÃO A CURTO PRAZO DE PREÇO DE ENERGIA UTILIZANDO MODELO HÍBRIDO COM SARIMAX E XGBOOST”

AUTOR: CÉZAR AUGUSTO FIGUEIREDO DA ROCHA


DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE SISTEMAS DE ENERGIA ELÉTRICA.

APROVADA EM: 15/05/2025


BANCA EXAMINADORA:

Documento assinado digitalmente
 MARIA EMÍLIA DE LIMA TOSTES
Data: 15/05/2025 15:48:07-0300
Verifique em <https://validar.iti.gov.br>


Prof.^a Dr.^a Maria Emília de Lima Tostes
(Orientadora – PPGEE/ITEC/UFPA)

Documento assinado digitalmente
 SUZANE CRUZ DE AQUINO MONTEIRO
Data: 15/05/2025 12:32:42-0300
Verifique em <https://validar.iti.gov.br>


Prof.^a Dr.^a Suzane Cruz de Aquino Monteiro

Documento assinado digitalmente
 UBIRATAN HOLANDA BEZERRA
Data: 15/05/2025 16:14:18-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Ubiratan Holanda Bezerra
(Avaliador Interno – PPGEE/ITEC/UFPA)

Documento assinado digitalmente
 FLAVIA PESSOA MONTEIRO
Data: 15/05/2025 12:26:09-0300
Verifique em <https://validar.iti.gov.br>

Prof.^a Dr.^a Flávia Pessoa Monteiro
(Avaliadora Externa – UFOPA)

Documento assinado digitalmente
 MARCIO JOSE MOUTINHO DA PONTE
Data: 15/05/2025 16:40:00-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Márcio José Moutinho da Ponte
(Avaliador Externo – UFOPA)

VISTO:

Prof. Dr. Diego Lisboa Cardoso
(Coordenador do PPGEE/ITEC/UFPA)

Dedico este trabalho a Deus, minha família e minha namorada Caroline, agradecendo pelo apoio, amor e inspiração durante esta jornada.

AGRADECIMENTOS

A Deus, por ser minha rocha e guia, sustentando-me nos momentos de incerteza.

À minha namorada, Caroline Cruz da Silva, minha companheira de sonhos e desafios, cujo carinho, paciência e incentivo me inspiraram a perseverar e acreditar em cada etapa desta jornada.

Aos meus colegas de quarto e amigos, Carlos e Paulo, pela amizade sincera, e por tudo que passamos juntos nessa caminhada..

À Prof. Dra. Maria Emília, minha orientadora visionária, por abrir as portas do Ceamazon, e me dar a oportunidade de realizar o mestrado.

À Prof. Dra. Flávia Monteiro, pela parceria de trabalho e amizade, por suas palavras encorajadoras, por acreditar em meu potencial e por tudo que fez por mim desde a graduação.

À Prof. Suzane Aquino, pela coorientação dedicada, pelo exemplo de dedicação acadêmica e por todos os momentos de cooperação nessa caminhada.

Ao Prof. Josivan Rodrigues, pelo suporte, pelas discussões inspiradoras e pela confiança depositada em meu trabalho.

Ao projeto P&D de Mercado de Energia Furnas, pela oportunidade de integrar pesquisa e prática, fornecendo dados, infraestrutura e a motivação necessária para transformar ideias em resultados.

Aos membros do laboratório, pelo espírito colaborativo, pelas trocas de conhecimento, pelas risadas compartilhadas nos momentos de tensão e pelo ambiente acolhedor que tornou este trabalho uma experiência inesquecível.

A todos que, de alguma forma, contribuíram para esta trajetória, meu mais profundo agradecimento e reconhecimento.

“A verdadeira arte da previsão está em filtrar o ruído do passado para revelar os contornos do futuro.”(SILVER, 2012, p. 45)

RESUMO

A previsão eficiente do preço de compra de energia elétrica a curto prazo é um desafio crucial, especialmente no cenário brasileiro, marcado pela alta volatilidade de preços e por um mercado energético dinâmico. As peculiaridades do setor elétrico nacional, como a dependência de fontes renováveis, variações climáticas sazonais e a crescente complexidade da integração de sistemas distribuídos, tornam essa tarefa ainda mais desafiadora. Grande parte dos métodos disponíveis foca na modelagem direta da série temporal, sem explorar plenamente a decomposição estrutural dos dados para aumentar a previsibilidade. Este trabalho propõe um modelo híbrido (SARIMAX-XGBoost) para previsão de curto prazo, combinando o modelo estatístico autorregressivo com médias móveis sazonais (SARIMAX) e o algoritmo de aprendizado extremo por reforço de gradiente (XGBoost). O SARIMAX captura componentes estruturais, como tendências, sazonalidade e efeitos de variáveis externas que estão intrínsecas no próprio histórico da variável alvo, enquanto o XGBoost modela resíduos e padrões complexos não explicados. A previsão final é obtida pela integração das saídas dos dois modelos. A validação foi realizada com dados reais do mercado brasileiro, abrangendo séries históricas de preços de energia elétrica. Os resultados experimentais mostram que o método proposto é eficaz em capturar a volatilidade do mercado, oferecendo previsões precisas e estáveis tanto em precisão quanto em estabilidade.

Palavras-chave: Previsão de preços de energia. Modelos híbridos. SARIMAX-XGBoost.

ABSTRACT

Efficient short-term forecasting of electricity purchase prices is a crucial challenge, particularly in the Brazilian context, characterized by high price volatility and a dynamic energy market. The peculiarities of the Brazilian electricity sector, such as dependence on renewable sources, seasonal climatic variations, and the growing complexity of integrating distributed systems, make this task even more demanding. Most available methods focus on direct time series modeling without fully leveraging the structural decomposition of data to enhance predictability. This study proposes a hybrid model (SARIMAX-XGBoost) for short-term forecasting, combining the seasonal autoregressive integrated moving average model (SARIMAX) with the extreme gradient boosting algorithm (XGBoost). SARIMAX captures structural components such as trends, seasonality, and external variable effects, while XGBoost models residuals and unexplained complex patterns. The final forecast is obtained by integrating the outputs of both models. Validation was performed using real market data from Brazil, covering historical series of electricity prices. Experimental results demonstrate that the SARIMAX-XGBoost model effectively captures market volatility, providing accurate and stable predictions, outperforming advanced methods such as artificial neural networks and purely XGBoost-based approaches in both accuracy and stability.

Keywords: Energy price forecasting. Hybrid models. SARIMAX-XGBoost.

LISTA DE FIGURAS

Figura 1 - Matriz Energética Brasileira 2023.....	11
Figura 2 – Produtos do mercado de energia elétrica e diferentes tempos de maturação.....	20
Figura 3 – Sistematização das metodologias de previsão do preço spot – electricity power forecasting.....	21
Figura 4 - Metodologia para estudo de anterioridade baseada na declaração PRISMA.....	22
Figura 5 - Publicações envolvendo electricity power forecast nos últimos anos.....	23
Figura 6 - Distribuição das publicações envolvendo electricity power forecast.....	23
Figura 7 - Mapa temático de palavras chaves separado por categorias relevantes.....	24
Figura 8 - Mapa temático de palavras chaves separado por ano de publicação.....	25
Figura 9 - Publicações envolvendo electricity forward curve nos últimos anos.....	26
Figura 10 - Distribuição das publicações envolvendo electricity forward curve.....	26
Figura 11- Mapa temático de palavras chaves separado por categorias relevantes.....	27
Figura 12- Mapa temático de palavras chaves separado por ano de publicação.....	28
Figura 13 - Fluxo de implementação do modelo híbrido SARIMAX-XGBoost para previsão de preços de energia.....	39
Figura 14 - Série histórica de preços Convencional Longo Prazo (LPC) de 2012 a 2024.....	41
Figura 15 - Decomposição da série temporal em componentes: série original, tendência, sazonalidade e resíduo.....	42
Figura 16 - Projeção da Sazonalidade para o ano de 2024.....	46
Figura 17 - Projeção da Tendência para o ano de 2024.....	47
Figura 18 - Importância das variáveis de entrada no modelo XGBoost.....	49
Figura 19 - Projeção do preço LPC para o ano de 2024.....	50
Figura 20 - Projeção do Resíduo para o ano de 2024.....	52
Figura 21 - Comparação entre os valores reais e previstos para o ano de 2024, considerando a recomposição completa.....	53

LISTA DE TABELAS

Tabela 1 – Tabela contribuições de trabalhos.....	37
Tabela 2 – Resultados da validação da projeção usando SARIMAX para Tendência e Sazonalidade.....	48
Tabela 3 – Estatísticas descritivas dos componentes estruturais.....	49
Tabela 4 – Resultados da validação da projeção usando XGBoost para o Resíduo.	52
Tabela 5 – Resultados da validação da projeção usando XGBoost para o Resíduo.	52
Tabela 6 – Resultados após a recomposição das projeções das componentes.....	54

SUMÁRIO

1 INTRODUÇÃO.....	11
1.1 Contextualização.....	11
1.2 Problemática.....	12
1.3 Objetivos.....	13
1.4 Justificativa.....	13
1.5 Estrutura da Dissertação.....	14
2. REFERENCIAL TEÓRICO.....	15
2.1 Revisão da Literatura sobre os Mercados de Energia.....	15
2.2 Ambientes de Contratação Livre (ACL) e Regulada (ACR).....	17
2.3 Formação de Preços.....	18
2.4 Tendências.....	19
2.5 Análise Sistemática da Bibliografia.....	20
2.5.2 Curvas Forward.....	25
2.6 Previsão de Preços de Energia Elétrica.....	28
2.7 Métodos Estatísticos Tradicionais.....	33
2.8 Métodos Baseados em Aprendizado de Máquina.....	35
2.9 Modelos Híbridos na Previsão de Séries Temporais.....	35
3. METODOLOGIA.....	37
3.1 Descrição do Modelo Híbrido SARIMAX-XGBoost.....	37
3.2 Fluxo de Implementação.....	38
3.4 Avaliação de Desempenho.....	44
4. RESULTADOS E DISCUSSÃO.....	45
4.1 Desempenho do Modelo XGBoost.....	49
4.2 Série Recomposta e Resultados Finais.....	52
5.1 Síntese dos Resultados.....	54
5.2 Limitações do Estudo.....	55
REFERÊNCIAS.....	57

1 INTRODUÇÃO

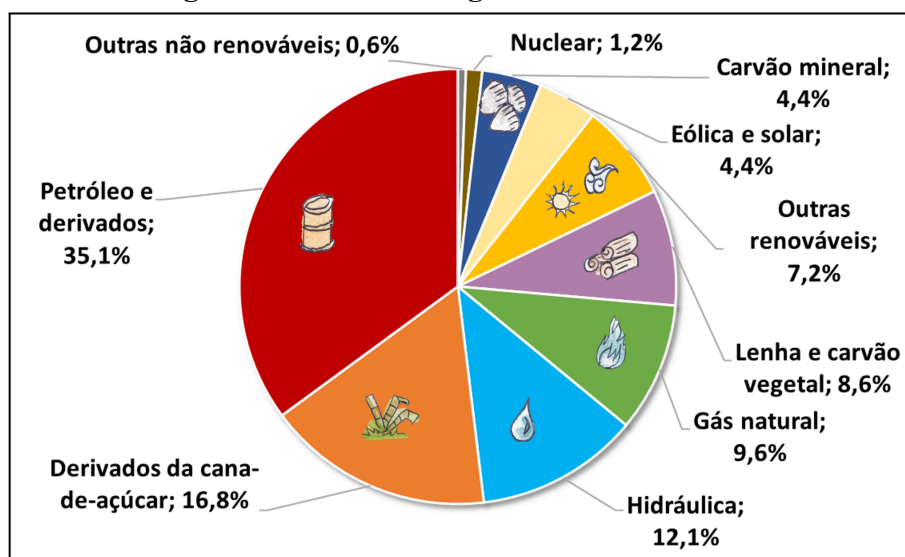
O setor elétrico brasileiro tem passado por transformações significativas nas últimas décadas, evoluindo de um modelo centralizado e predominantemente estatal para um ambiente mais dinâmico e competitivo. Nesse contexto, a previsão eficiente do preço de compra de energia elétrica a curto prazo emerge como um desafio crucial para diversos agentes do mercado, incluindo comercializadoras, grandes consumidores e órgãos reguladores. A volatilidade característica desse mercado, influenciada por fatores como condições climáticas, disponibilidade de recursos hídricos, demanda energética e políticas regulatórias, torna essa tarefa particularmente complexa e desafiadora.

1.1 Contextualização

A previsão precisa do preço de compra de energia elétrica desempenha um papel essencial na estabilidade e eficiência do setor elétrico, impactando diretamente a tomada de decisões por consumidores, agentes do mercado e órgãos reguladores. No Brasil, onde a matriz energética é predominantemente renovável e altamente influenciada por variáveis climáticas, prever com precisão o comportamento do mercado energético é um desafio crítico.

A volatilidade dos preços resulta de múltiplos fatores, incluindo a sazonalidade das chuvas, oscilações na demanda, políticas tarifárias, custos de geração e variações na oferta de fontes intermitentes, como eólica e solar, a **Figura 1** ilustra a composição da matriz energética brasileira em 2023, destacando a participação significativa das fontes renováveis, como a hidrelétrica, que historicamente responde por grande parte da geração elétrica no país. Esse perfil energético reforça que a disponibilidade dessas fontes está sujeita a condições climáticas e variações sazonais.

Figura 1 - Matriz Energética Brasileira 2023.



Fonte: Empresa de Pesquisa Energética (EPE), 2025.

Nos últimos anos, a crescente digitalização do setor elétrico e o aumento da disponibilidade de dados históricos estimularam o desenvolvimento de abordagens mais sofisticadas para previsão de preços de energia elétrica. Métodos estatísticos tradicionais, como ARIMA e SARIMA, continuam sendo amplamente utilizados para modelagem de séries temporais devido à sua capacidade de capturar tendências e sazonalidade. No entanto, essas abordagens apresentam limitações na modelagem de padrões não lineares e dinâmicas complexas, aspectos fundamentais em mercados energéticos cada vez mais dinâmicos e interdependentes (BOX et al., 2015). Por outro lado, métodos de aprendizado de máquina, como redes neurais profundas e algoritmos de boosting, vêm se destacando por sua capacidade de identificar padrões não lineares nos dados, embora muitas vezes desconsiderem informações estruturais relevantes, como tendências e sazonalidade (CHEN e GUESTRIN, 2016; HOCHREITER e SCHMIDHUBER, 1997).

Nesse cenário, modelos híbridos que combinam métodos estatísticos e técnicas de aprendizado de máquina surgem como uma alternativa promissora para melhorar a precisão da previsão de preços de energia elétrica. A proposta deste estudo é explorar essa abordagem híbrida por meio do modelo SARIMAX-XGBoost, que integra a robustez de um modelo estatístico para capturar tendências e sazonalidade com a flexibilidade do aprendizado de máquina para modelar padrões residuais mais complexos.

1.2 Problemática

A previsão de preços de energia elétrica no contexto brasileiro enfrenta desafios específicos que limitam a eficácia dos métodos convencionais. Grande parte dos modelos disponíveis foca na modelagem direta da série temporal, sem explorar plenamente a decomposição estrutural dos dados para aumentar a previsibilidade. Essa abordagem simplificada frequentemente não consegue capturar adequadamente a complexidade multifatorial que influencia os preços de energia.

Os modelos estatísticos tradicionais, embora ofereçam boa interpretabilidade, apresentam limitações significativas ao lidar com dados não estacionários e relações não lineares. Por outro lado, técnicas avançadas de aprendizado de máquina, como redes neurais e algoritmos baseados em árvores de decisão, podem modelar relações complexas, mas frequentemente funcionam como "caixas-pretas", dificultando a compreensão dos fatores que influenciam suas previsões.

Diante desse cenário, emerge a necessidade de desenvolver abordagens híbridas que possam combinar as vantagens de diferentes métodos, aproveitando a interpretabilidade dos modelos estatísticos e o poder preditivo dos algoritmos de aprendizado de máquina. A questão central que este trabalho busca responder é: como desenvolver um modelo híbrido que integre eficientemente técnicas estatísticas e de aprendizado de máquina para melhorar a precisão na previsão de preços de energia elétrica a curto prazo no mercado brasileiro?

1.3 Objetivos

1.3.1 Objetivo Geral

Desenvolver e validar um modelo híbrido que combine o modelo estatístico autorregressivo com médias móveis sazonais (SARIMAX) e o algoritmo de aprendizado extremo por reforço de gradiente (XGBoost) para previsão a curto prazo de preços de energia elétrica no mercado brasileiro, visando superar as limitações dos métodos individuais e melhorar a precisão preditiva.

1.3.2 Objetivos Específicos

1. Analisar as características específicas das séries temporais de preços de energia elétrica no mercado brasileiro, identificando padrões de sazonalidade, tendência e componentes residuais;
2. Implementar e otimizar um modelo SARIMAX para capturar os componentes estruturais da série temporal, incluindo tendências, sazonalidade e efeitos de variáveis externas;
3. Desenvolver um modelo XGBoost para modelar os resíduos e padrões complexos não explicados pelo componente estatístico;
4. Integrar os modelos SARIMAX e XGBoost em uma arquitetura híbrida que maximize o aproveitamento das vantagens de cada abordagem;
5. Avaliar o desempenho do modelo híbrido proposto em comparação com métodos tradicionais e técnicas isoladas, utilizando métricas de avaliação robustas;
6. Analisar a importância relativa das variáveis preditoras e dos componentes estruturais na formação dos preços de energia elétrica.

1.4 Justificativa

A abordagem híbrida proposta neste trabalho se justifica pela necessidade de melhorar a capacidade preditiva dos modelos tradicionais, que, isoladamente, apresentam limitações ao lidar com dados voláteis e não lineares. A combinação de técnicas estatísticas com aprendizado de máquina permite explorar o melhor de cada abordagem: o SARIMAX captura a estrutura subjacente da série temporal, enquanto o XGBoost modela os padrões residuais complexos, garantindo previsões mais precisas e robustas.

Do ponto de vista teórico, este estudo contribui para o avanço do conhecimento na área de previsão de séries temporais, especialmente no contexto de mercados energéticos com alta volatilidade. A integração de modelos estatísticos e de aprendizado de máquina representa uma evolução natural na busca por métodos mais eficazes para lidar com a complexidade crescente dos dados.

Além do impacto acadêmico, este estudo tem implicações práticas significativas para o setor energético brasileiro. Modelos mais precisos podem auxiliar comercializadoras de energia,

grandes consumidores e órgãos reguladores a mitigar riscos financeiros e melhorar a gestão da compra de energia elétrica. Em um mercado onde pequenas variações percentuais podem representar milhões de reais, a melhoria na precisão das previsões tem impacto direto na competitividade e sustentabilidade financeira dos agentes.

Dessa forma, a proposta deste estudo não apenas contribui para o avanço do conhecimento científico na área de previsão de séries temporais, mas também oferece uma ferramenta prática para aprimorar a tomada de decisão no mercado energético brasileiro, com potencial para gerar benefícios econômicos tangíveis para diversos stakeholders.

1.5 Estrutura da Dissertação

A dissertação está organizada em cinco capítulos principais:

- **Capítulo 1 - Introdução:** apresenta o contexto do estudo, a problemática abordada, os objetivos gerais e específicos, a justificativa do trabalho e a estrutura da dissertação.
- **Capítulo 2 - Referencial Teórico:** discute os principais conceitos e técnicas utilizados na previsão de preços de energia elétrica, abordando modelos estatísticos tradicionais, aprendizado de máquina e abordagens híbridas, além de contextualizar o cenário brasileiro no setor energético.
- **Capítulo 3 - Metodologia:** detalha a abordagem utilizada para a construção do modelo híbrido SARIMAX-XGBoost, incluindo o fluxo de implementação, a descrição do conjunto de dados e as métricas utilizadas para avaliação do desempenho preditivo.
- **Capítulo 4 - Resultados e Discussão:** apresenta os experimentos realizados e analisa os resultados obtidos, comparando o desempenho do modelo híbrido com métodos tradicionais e isolados de aprendizado de máquina.
- **Capítulo 5 - Conclusão e Trabalhos Futuros:** sintetiza os principais achados da pesquisa, discute as limitações do estudo e sugere possíveis direções para futuras investigações na área.

2. REFERENCIAL TEÓRICO

2.1 Revisão da Literatura sobre os Mercados de Energia

Tomando o cenário mundial em perspectiva, observa-se que diferentes países adotam distintos desenhos para seus mercados de energia que refletem intrinsecamente questões históricas, culturais, regulatórias e de disponibilidade das fontes primárias. Entretanto, um ponto comum a ser destacado é o impacto dos desenhos de mercado nos preços verificados ao consumidor, e na complexidade e sucesso de ferramentas para previsão dos preços da energia. Tal compreensão é de vital importância, pois a previsão de preços da energia elétrica em um determinado mercado está intrinsecamente ligada à configuração do mercado de energia para o ambiente em que esta será realizada, refletindo o impacto regulador do Estado e as forças de mercado de oferta e demanda entre os comercializadores e consumidores.

De uma forma geral, os modelos de desenho de mercado observados a nível mundial se relacionam diretamente com o grau de participação do consumidor na definição do supridor da energia elétrica, mesmo em mercados altamente competitivos, permanecendo atrelados ao modelo de monopólios naturais. Desta forma, os modelos de mercado atualmente observados se dividem em (VIANA, 2018):

- Monopólio;
- Single Buyer;
- Competição no Atacado
- Competição no Varejo.

2.1.1 Monopólio

Trata-se do modelo verificado em boa parte do século XX, até a implementação das reformas estruturantes verificadas nos mercados de energia de diversos países. Neste modelo, todas as funções da indústria são realizadas de forma verticalizada (abrangendo geração, transmissão e distribuição) por uma única empresa, visando a obtenção de ganhos de sinergia e facilitação do controle do produto estratégico a nível nacional.

No cenário brasileiro, tal modelo foi observado até meados dos anos 2000, quando da implementação de reformas de modernização do setor elétrico após cenários de escassez e apagões ocorridos no início da década. Até então prevaleciam dois modelos de monopólio: uma variante na qual uma única estatal atuava realizando todas as funções da indústria e outra na qual uma empresa federal atuava na geração e transmissão, enquanto uma distribuidora estadual atuava no setor de distribuição de energia (VIANA, 2018).

2.1.2 Single Buyer

O modelo de comprador único surgiu nos Estados Unidos na década de 1970 visando incrementar a eficiência alocativa dos recursos diante do cenário de crise energética resultante do choque do petróleo. Neste modelo, surge a figura do IPP (Independent Power Producer) que se constitui em agentes privados que competem para fornecer energia elétrica para os centros de carga já existentes por meio da participação em leilões ou chamadas públicas reguladas por uma agência governamental.

Tal modelo trata-se de um primeiro estágio no sentido da implementação de um mercado competitivo, pois, apesar de haver competição dos agentes no momento da ocorrência dos leilões, ainda assim os consumidores não dispõem de escolha do seu supridor de energia (VIANA, 2018).

2.1.3 Competição no Atacado

Tal modelo teve início em sua implementação a nível mundial na década de 1980 e posterior consolidação na década de 1990. Neste modelo, as regulamentações por custo e comercialização da energia elétrica são suprimidas, permitindo que os grandes consumidores e distribuidoras comprem de forma competitiva dos agentes geradores e comercializadores. As distribuidoras de energia ainda permanecem com a tarefa de entregar energia fisicamente devido ao monopólio natural da área de concessão e a representar comercialmente os pequenos consumidores.

Neste modelo surgem dois novos atores: o operador de mercado, responsável pela contabilização e liquidação financeira das posições; e o operador de sistema, responsável pelo equilíbrio e manutenção da estabilidade física do consumo e geração. Tal modelo apresenta predominância em sua adoção pelo setor elétrico brasileiro (VIANA, 2018).

2.1.4 Competição no Varejo

Este modelo é caracterizado por uma participação ativa dos consumidores nas operações do mercado, independentemente do seu tamanho ou aspectos técnicos tais como nível de tensão e potência instalada para ingresso. Entretanto, ainda podem existir agentes como comercializadores varejistas ou agregadores que negociam diretamente com uma base de clientes, mediante a compra de energia no mercado atacadista, assumindo riscos de mercado pela sua atuação.

Para que este modelo funcione adequadamente, é necessário um grande número de comercializadores e consumidores e que estes detenham ferramentas de comparação de preços e realização de transações, conferindo liquidez ao mercado.

Neste cenário, o poder estatal centralizado na formação dos preços é bastante reduzido, fazendo com que os mecanismos de mercado tais como a oferta e demanda sejam mais determinantes na definição dos preços da energia elétrica. Neste ambiente,

semelhantermente a uma bolsa de valores, o preço da energia elétrica estará atrelado mais diretamente à percepção do valor do produto e assunção de riscos do que necessariamente à estrutura de custo de geração e disponibilização (VIANA, 2018).

2.2 Ambientes de Contratação Livre (ACL) e Regulada (ACR)

No mercado brasileiro a partir da publicação da Lei Nº 10.848/2004, as transações bilaterais de compra e venda de energia são realizadas em dois ambientes: Ambiente de Contratação Livre (ACL), cujos contratos são de livre negociação entre os agentes envolvidos, respeitada a regulamentação vigente emitida pela Câmara de Comercialização de Energia (CCEE); e Ambiente de Contratação Regulada (ACR), subdivididos nos Contratos de Comercialização de Energia no Ambiente Regulado (CCEAR), Contratos de Cota de Garantia Física (CCGF), Contratos de Cotas de Energia Nuclear (CCEN), Contratos de Itaipu, Contratos do PROINFA, Contratos de Leilão de Ajuste e Contratos Bilaterais Regulados (CCEE, 2023a).

De uma forma geral, o ACR visa atender a demanda dos consumidores cativos, que são os consumidores residenciais, de serviço e indústrias de menor porte. No ACR as Distribuidoras de energia elétrica atendem os consumidores cativos através de uma contratação em um ambiente pool, realizada por meio de leilões de energia elétrica, nos quais os geradores que oferecem os menores preços são os ganhadores. Por meio dos CCEARs, as distribuidoras de energia são obrigadas a contratar 100% de sua carga de energia a partir da contratação de lastro de energia e potência. Os termos e as condições dos CCEARs e os mecanismo de garantias bilaterais para assegurar o cumprimento das obrigações contratuais são definidos pela ANEEL, no âmbito de sua competência para elaboração dos editais dos leilões de energia, conforme previsto na Lei nº 10.848/2004 e pelo Decreto nº 5.163/2004 (CASTRO & BRANDÃO, 2021).

Este ambiente é caracterizado por uma sistemática de leilões organizados pelo governo, nos quais são firmados contratos bilaterais de longo prazo entre geradores e distribuidoras de energia, em que estas arcam com o risco de crédito dos consumidores cativos. Há leilões de vários tipos, dentre os quais se destacam os Leilões de Energia Existente, para compra de energia de usinas já em operação comercial, e os Leilões de Energia Nova, destinados a viabilizar a construção de novos empreendimentos de geração, sendo estes realizados majoritariamente no atual cenário, com contratos de longo prazo, de até 30 anos (CASTRO & BRANDÃO, 2021).

Por outro, o ACL é composto de consumidores de maior volume de consumo, com a comercialização de energia realizada de forma bilateral entre geradores, comercializadores, consumidores livres e especiais. Neste ambiente, os termos de contratação como preço, quantidade, condições de registro dos contratos, garantias, dentre outros, são pactuados livremente entre os agentes. Desta forma, a CCEE se limita a registrar os prazos e as quantidades destes contratos, sendo seus valores restritos às partes contratantes (CASTRO et. al, 2017a).

O ACL está estruturado como um balcão não organizado, sem uma entidade central para administrar os contratos bilaterais firmados. Apesar disso, ocorreram medidas para a criação de sistemas de negociação de contratos padronizados por meio de entidades privadas, sendo a mais bem sucedida o Balcão Brasileiro de Comercialização de Energia (BBCE) (CASTRO et. al, 2017b). Entretanto, o volume de contratos firmados nesta plataforma ainda se mostra incipiente se comparado aos contratos firmados diretamente entre comercializadores e consumidores e que não possuem seus preços divulgados o que, conseqüentemente, prejudica o conhecimento dos agentes envolvidos dos preços usualmente praticados no mercado, representando uma oportunidade para o desenvolvimento de metodologias e ferramentas para a previsão de preços de comercialização.

As tendências para a comercialização de energia para o mercado brasileiro nos próximos anos é uma acentuação do crescimento do ACL e retração do ACR, com migração definitiva da maioria das empresas (BRANDÃO, 2019). Segundo o Anuário Estatístico de Energia Elétrica (EPE, 2023), com dados relativos ao ano de 2022, o percentual de consumo no ACL já representa 39,66% do total de energia consumida no Brasil, o que se traduz na entrada contínua de novos agentes e mais liquidez nos contratos firmados.

2.3 Formação de Preços

O mercado brasileiro de energia elétrica apresenta características únicas se comparado aos demais mercados internacionais em função do desenho contratual com mecanismos que garantem a expansão equilibrada e dinâmica entre a oferta e a demanda em um mercado competitivo com baixa participação da geração térmica (CASTRO et. al, 2014). Desta forma, o mercado de energia trata-se de um mercado de contratos e não de energia, de longo prazo em que os agentes consumidores necessitam garantir 100% de contratação do seu consumo, sob pena de sujeição a penalidades.

Adicionalmente, ao contrário dos mercados europeus e estadunidense, o despacho é realizado de forma centralizada, com base em um modelo computacional de otimização de custos e não a partir de ofertas de preços e quantidades dos geradores e consumidores. Tal desenho se deve, sobretudo, às características específicas do mercado brasileiro, marcado pela predominância da geração hídrica. Desta forma, o preço de curto prazo da energia, definido com base no custo marginal de operação calculado pelos modelos computacionais NEWAVE, DECOMP e DESSEM tende a ser demasiado baixo sempre que ocorrerem condições hidrológicas normais e demasiado elevado em caso de condições hidrológicas desfavoráveis. Em essência, portanto, os valores assumidos pelo Preço da Liquidação das Diferenças (PLD) são resultantes da hidrologia e não da interação entre preferências dos consumidores e estruturas de custos dos geradores (CASTRO et. al, 2014). Tal característica se torna evidente em momentos de crise hídrica com a observada no ano de 2021, em que os reservatórios do Sudeste e Centro-Oeste, responsáveis por 70% do armazenamento de energia do país, estavam no patamar mínimo histórico de 26% (SANTOS & FIGER, 2022).

Devido a tal volatilidade, o sistema brasileiro não possui uma das principais características de um mercado competitivo, que é a capacidade de autorregulação através de

ações dos agentes participantes induzidos pelos preços de mercado. Isto se deve ao fato de o preço de curto prazo não fornecer uma sinalização adequada para o aumento ou diminuição da oferta, para novos investimentos ou para desinvestimentos no setor: um preço de curto prazo elevado não é por si só indicador de necessidade de investimentos e um preço muito baixo não é indicação clara de que há excesso de capacidade ociosa (CASTRO et. al, 2014). Desta forma, há uma discrepância acentuada entre o PLD e os custos de produção da geração de energia elétrica. E esta discrepância não dá sinais econômicos para orientar as decisões de investimento e desinvestimento por parte dos geradores (CASTRO et. al, 2017B).

Portanto, o PLD não fornece sinal econômico para investimentos e desinvestimentos para os agentes do mercado de energia. Como consequência direta, ele não pode servir de base para a expansão do sistema, pois é imprevisível no curto, no médio e no longo prazos e assim constitui base inadequada para a tomada de decisões de investimento e financiamento (CASTRO et. al, 2017B), bem como transações de blocos de energia livremente negociados entre os agentes do mercado.

Como evidência dessa disparidade, no mercado brasileiro as atividades comerciais e físicas são realizadas por agentes distintos: ao passo que a Câmara de Comercialização de Energia Elétrica (CCEE) é o operador comercial que registra as transações de compra e venda de energia e é responsável pela liquidação financeira do mercado de curto prazo de energia, o Operador Nacional do Sistema Elétrico (ONS) é responsável pela coordenação e controle da operação das instalações de geração e transmissão de energia elétrica (PORTO & RODRIGO, 2019).

Desta forma, a CCEE calcula o PLD com capilaridade horária para cada patamar de carga e submercado, com limitações em um preço máximo e um preço mínimo, vigentes para cada período de apuração. O cálculo do PLD é baseado no despacho *ex-ante*, ou seja, é apurado com as informações previstas, anteriores à operação real do sistema, considerando os valores de disponibilidades declaradas de geração e a demanda prevista para cada submercado. Na prática, adota-se uma abordagem *tight pool*, conhecida como preço por custo, sem a realização dos leilões do tipo *day-ahead* ou *intra-day* para a operação do sistema (VIANA, 2018).

Diante deste paradigma, tanto os agentes participantes do ACL quanto do ACR dependem de metodologias e ferramentas para a precificação da energia sob os diversos horizontes de curto, médio e longo prazos. Tais metodologias devem ser capazes de refletir as interações entre oferta e demanda e o prêmio do risco envolvido na comercialização da energia, trazendo menos volatilidade ao setor e permitindo a viabilidade econômica de novos empreendimentos apesar de períodos extremos como os observados em casos de crise hídrica.

2.4 Tendências

No médio e longo prazo, o Sistema Elétrico Brasileiro caminha para um gradativo processo de liberalização da compra e venda de energia, com predominância do mercado livre

frente ao mercado cativo, tendência já consolidada de forma acentuada nos Estados Unidos e na União Europeia (CASTRO et. al, 2020).

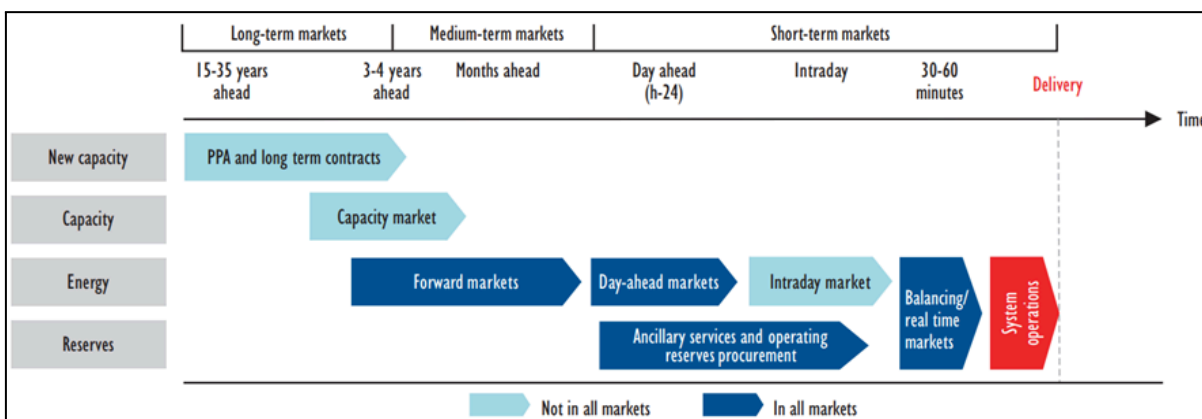
Neste novo paradigma, as distribuidoras de energia de energia irão atuar como responsáveis pelos investimentos, operação e manutenção da rede de distribuição, mantendo a estrutura de monopólio natural, sendo remuneradas através da tarifa de uso da rede (Tarifa de Uso do Sistema de Distribuição – TUSD, no caso brasileiro). A comercialização da energia elétrica, por sua vez, é tratada como uma atividade competitiva, desvinculada das distribuidoras, atuando diretamente no mercado livre (CASTRO et. al, 2020).

Tal configuração irá fortalecer, certamente, plataformas centralizadas de comercialização, tais como a BBCE, que permitam que os agentes envolvidos transacionem os montantes de energia com diferentes escalas de tempo, de acordo com os riscos assumidos e disposição para exposição ao mercado de curto prazo (BRANDÃO, 2023). Neste cenário, de maior liquidez dos contratos e operação centralizada, a precificação da energia sob os diversos horizontes torna-se ainda mais relevante pois pode significar oportunidades de ganhos para os agentes envolvidos, bem como risco de perda para aqueles subcontratados.

2.5 Análise Sistemática da Bibliografia

É, então, importante ressaltar os diferentes produtos e suas respectivas escalas de tempo no ambiente do mercado de eletricidade, conforme apresentado na **Figura 2**, para identificar as áreas de maior interesse nesse tópico. Portanto, no longo prazo (*long-term markets*), estão usualmente presentes as estruturas relacionadas aos contratos de disponibilidade de novas unidades de geração (PPAs – *Power Purchase Agreements*). No médio prazo são firmados contratos de capacidade para as usinas já existentes, além de contratos de entrega de energia futura (contratos *forward*). No curto prazo, por sua vez, estão localizadas as operações do mercado de energia e de serviços ancilares nos horizontes *Day-Ahead Market* (DAM) e o *Intraday/Real Time Market* (IDM ou RTM).

Figura 2 – Produtos do mercado de energia elétrica e diferentes tempos de maturação



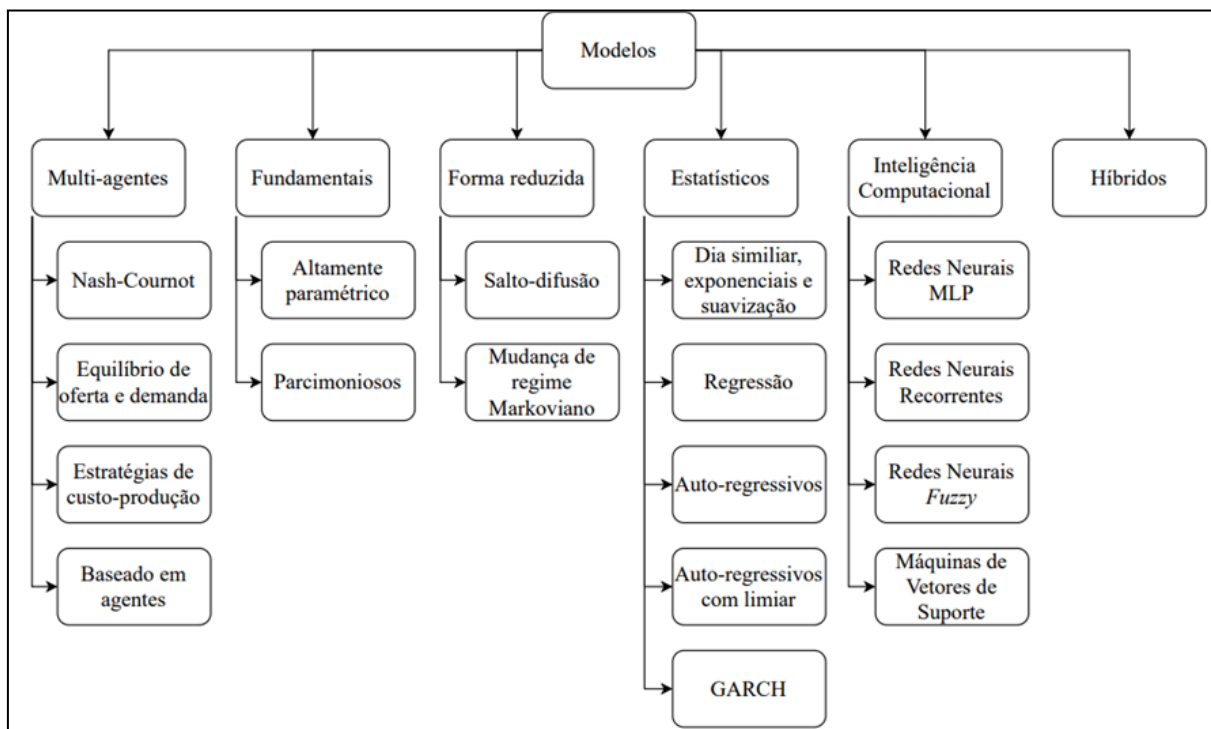
Fonte: IEA, 2016.

Assim, considerando os diferentes produtos e diferentes horizonte de tempo de maturação, tem-se o estabelecimento das diferenças fundamentais entre os dois conceitos envolvidos na previsão de preços: *Electricity Power Forecasting* e *Curvas Forward*.

2.5.1 Electricity Power Forecasting

Electricity Power Forecasting trata-se essencialmente da previsão do preço *spot*, ou do mercado de curto prazo, da energia em um determinado instante de tempo. No mercado brasileiro, é resultado da cadeia de resultados dos modelos do NEWAVE, DECOMP e DESSEM, que estão intrinsecamente atrelados aos fundamentos operacionais e técnicos do sistema. A nível mundial, a literatura reporta uma diversidade de metodologias e técnicas para definição do preço, cujo extrato está apresentado na **Figura 3**: multi-agentes; fundamentais; forma reduzida; estatísticos; inteligência computacional; híbridos.

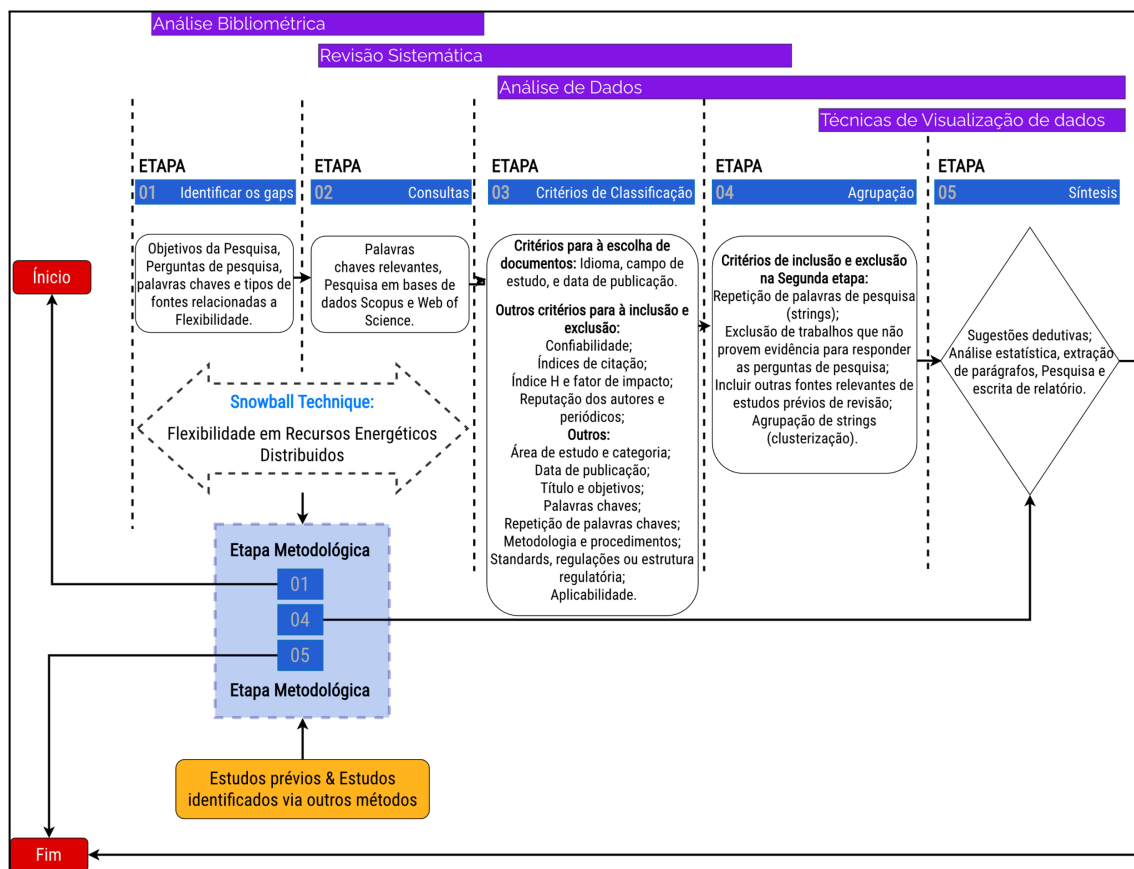
Figura 3 – Sistematização das metodologias de previsão do preço spot – electricity power forecasting



Fonte: NAMATELA, 2023.

Visando implementar uma análise sistemática da bibliografia relativa ao tema, foi desenvolvido um estudo de anterioridade baseado na declaração PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), desenvolvido por um grupo de acadêmicos de saúde pública conforme apresentado em (MOHER et al., 2009). A metodologia dividida em 5 (cinco) etapas, apresentada na **Figura 4**, inclui uma análise bibliográfica que visa mapear a produção científica, identificar padrões de pesquisa e avaliar a influência de periódicos e instituições na área de em questão.

Figura 4 - Metodologia para estudo de anterioridade baseada na declaração PRISMA

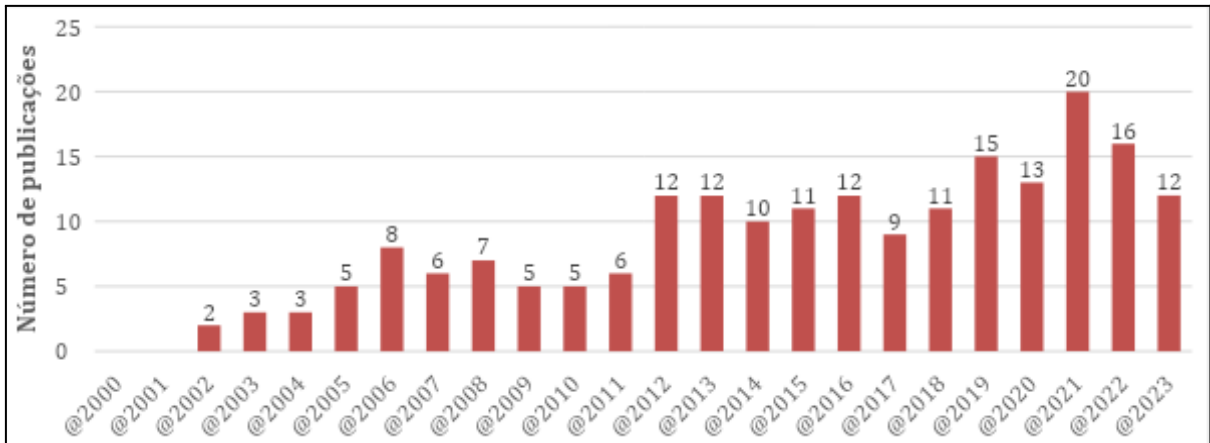


Fonte: Autores, 2025.

Portanto, visando a realização desse levantamento, foi utilizada a plataforma Scopus para seleção dos trabalhos objeto da análise. Tal plataforma se trata da maior base de dados de resumos e citações de literatura técnica-científica revisada por pares, com capacidade de indexação dos mais relevantes repositórios, além de possuir ferramentas bibliográficas para acompanhar, analisar e visualizar a pesquisa.

Desta forma, um levantamento das publicações contendo o termo "*electricity*" *power forecast* foi realizado considerando os últimos 23 anos, conforme apresentado na **Figura 5**. Os resultados mostram que a partir do ano de 2014 as publicações aumentaram significativamente, mantendo o ritmo de crescimento no ano de 2023 ainda corrente.

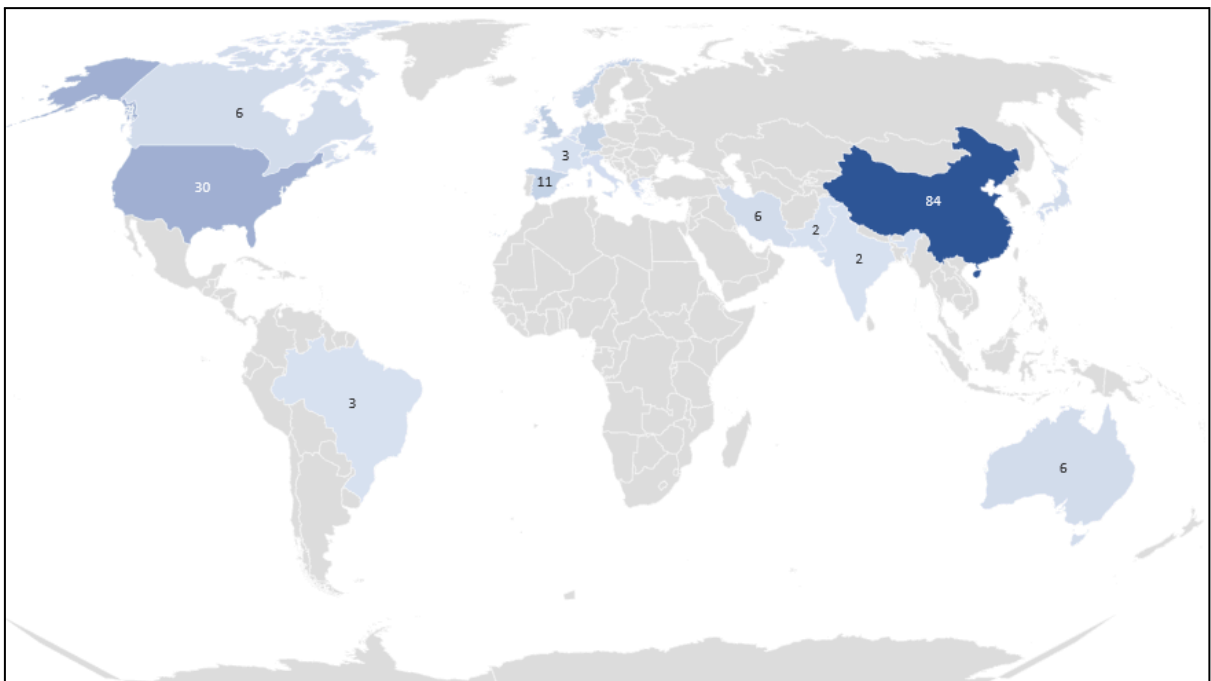
Figura 9 - Publicações envolvendo *electricity forward curve* nos últimos anos.



Fonte: Autores, 2025.

Mundialmente, as pesquisas envolvendo o *electricity power forecast* são lideradas pela China (84), seguida pelos Estados Unidos (30) e Reino Unido (14). O Brasil aparece na 17ª posição, com um total de 3 publicações, conforme **Figura 10**.

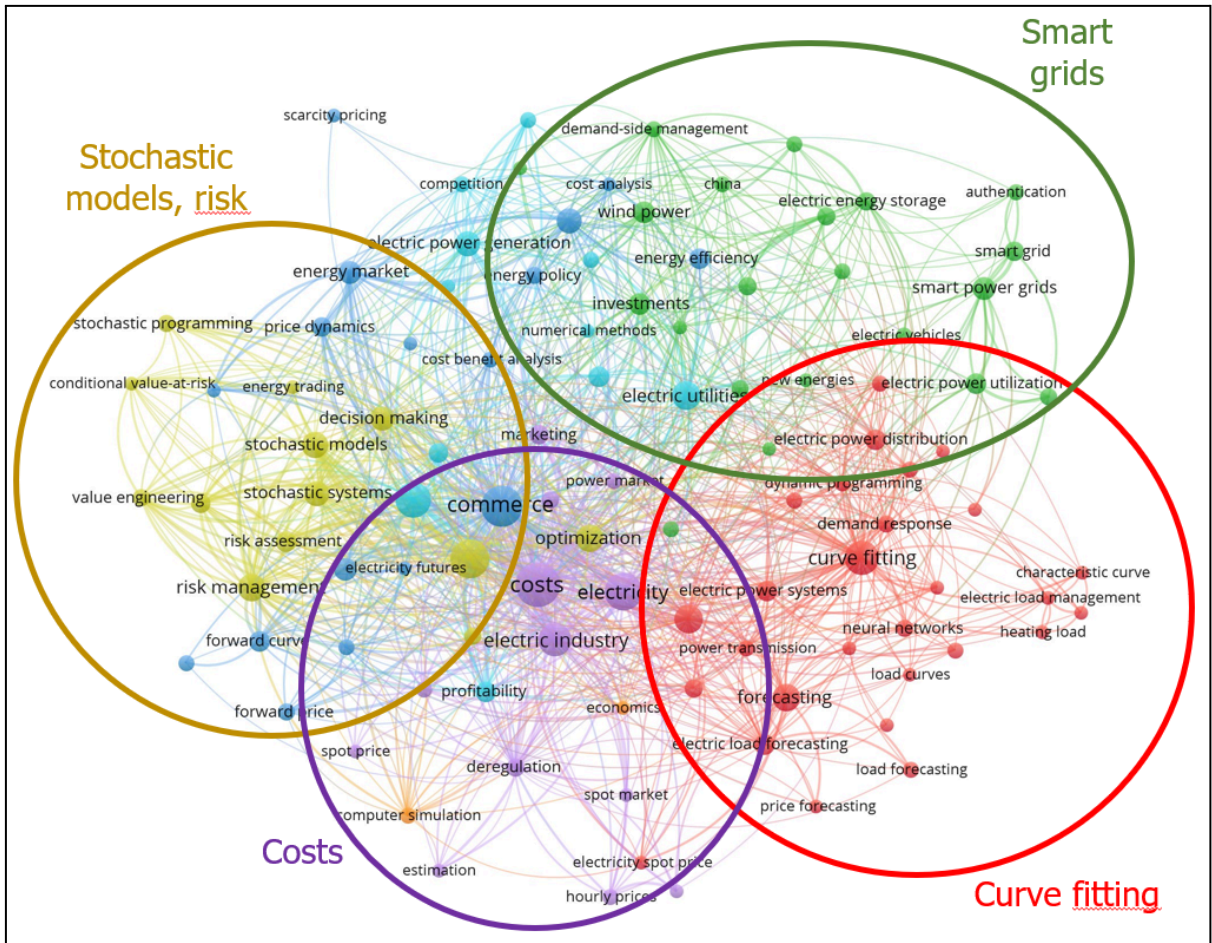
Figura 10 - Distribuição das publicações envolvendo *electricity forward curve*.



Fonte: Autores, 2025.

A **Figura 11** apresenta o resultado da clusterização da relação das palavras chave das publicações da base de dados para a *electricity forward curve*, ressaltando as relações entre as palavras-chave observadas em cada cluster. Dado o mapa apresentado, cabe ressaltar dois grupos principais: técnicas de modelagem da curva e análise de risco utilizando a ferramenta.

Figura 11- Mapa temático de palavras chaves separado por categorias relevantes.



Fonte: Autores, 2025.

De forma a rastrear os tópicos mais relevantes na atualidade na base de dados selecionada, o software permite realizar o mapeamento das palavras chave identificadas conforme o ano das publicações em que estão associadas, conforme apresentado na **Figura 12**. Conforme pode ser observado, as publicações atualmente mais relevantes se concentram nas áreas de inteligência computacional, envolvendo técnicas como machine learning, deep learning, árvores de decisão, random forests e redes neurais.

Tal desacoplamento é verificado com mais intensidade nos horizontes de médio e longo prazo, que estão menos suscetíveis a variáveis tipicamente operacionais e mais suscetíveis a variáveis que modelam o risco associado ao negócio (BARROSO et. al., 2007). Desta forma, mediante tal desacoplamento nos horizontes de médio e longo prazo, faz-se necessário o estabelecimento de metodologias que tornem projeções de curvas de preços que refletem não somente aspectos intrínsecos à operação da rede, mas também o risco associado à comercialização, de forma a otimizar a modelagem dos contratos firmados entre os agentes.

Nesse contexto, introduz-se o conceito de curvas *forward* como ferramentas capazes de capturar as expectativas mercado tais como prêmio de risco, custo de oportunidade, liquidez, concentração de mercado e volatilidade de forma a refletir com maior acurácia o preço de comercialização da energia (PILOPOVIC, 1998).

Vários estudos têm sido realizados ao longo dos últimos anos com o objetivo de compreender e modelar as curvas *forward* de preço de energia elétrica nos horizontes de médio e longo prazo. Um dos trabalhos seminais na área trata-se do estudo realizado por (BESSEMBINDER & LEMON 2002), o qual propõe um modelo de equilíbrio para o mercado de energia, permitindo analisar o sistema através de um número definido de agentes sob análise de oferta. Investigaram o mercado americano de energia e suas taxas *spot* e *forward* através de um modelo discreto no tempo e identificaram que o prêmio de risco no mercado futuro depende variação e assimetria de cauda direita dos preços *spot*. Identificaram também um prêmio de risco decrescente na curva *forward* com a entrada de especuladores no mercado.

Nesse mesmo sentido, foi proposta a construção de uma curva de preço *forward* da energia no mercado comum do norte europeu utilizando como base os preços de contratos a termo disponíveis para diferentes datas no mesmo mercado (FLETEN & FLEMING, 2003). Os autores relatam que, mesmo para sistemas elétricos com despacho centralizado, como é o caso do mercado norte europeu, é necessário inserir variáveis que reflitam a aversão ao risco por parte dos agentes de mercado, necessariamente atrelando a construção das curvas *forward* a modelos financeiros.

Observando o mercado brasileiro, o trabalho realizado em (RODOLFO et. al., 2015) se propôs a realizar uma análise da situação da gestão de risco de mercado na comercialização de energia elétrica no país por meio de uma pesquisa com os principais agentes atuantes no setor elétrico, incluindo geradores, comercializadores e consumidores. Por meio da pesquisa foi verificado que as empresas que possuem maior aversão ao risco são aquelas com perfil mais voltado para geração, ao passo que as comercializadoras apresentam, de uma forma geral, maior exposição ao risco. Porém, o setor como um todo tem aprimorado suas metodologias pois, as deficiências na gestão do risco têm sido traduzidas em perdas financeiras. Foi verificado na pesquisa que, ainda que muitas empresas ainda utilizam o PLD como insumo de gestão de risco, tem sido crescente o uso das curvas *forward* como estratégia de implementação do valor em risco na comercialização.

O trabalho apresentado em (LEMEA et. al., 2015) detalhou as particularidades do mercado nacional e utilizou-se de informações de um serviço de fornecimento de preços de

fechamento de contratos individuais no ambiente de contratação livre (ACL) com o objetivo de criar uma curva de consenso de preços e analisar suas características. Foi observado que os preços de vencimento mais curtos (dentro do ano corrente) apresentaram características diferentes dos preços dos vencimentos mais longos da curva, com variações até cinco vezes maiores, indicando menor risco associado a posições de longo prazo.

Tomando novas metodologias de abordagem da formação da curva de preços, pode-se citar o trabalho realizado em (YOUSEFI et. al., 2019) o qual focou no emprego de técnicas de *machine learning* para a previsão de preços de energia mensalmente em um horizonte de cinco anos para o estado da Califórnia. Para isso, foram coletados dados históricos ao longo de dezesseis anos, totalizando setenta e oito variáveis analisadas com o objetivo de determinar a correlação entre as mesmas e a sua influência nos preços de médio e longo prazo.

Além dessas, em (MENDONÇA et. al., 2019) foi realizado um teste comparativo entre quatro modelos preditivos: SARIMAX (Seasonal Autoregressive Integrated Moving Average), LSTM (Long Short Term Memory), GRU (Gated Recurrent Unit) e CNN-LSTM (Convolutional Neural Networks). Desta forma, foram realizadas previsões semanais sobre as séries de preços futuros de maturidade de 0 a 4. Em todas as séries realizadas, foi observado que os algoritmos GRU e SARIMAX demonstraram precisão superior em relação aos demais testados.

Em (PAPPAS & SANTOS, 2019) foi avaliado o poder de mercado na formação de preços via oferta, visando analisar a influência de fatores como a quantidade de agentes e a presença de contratos bilaterais. Tal estudo visou, sobretudo suscitar a discussão sobre a mudança de paradigma da formação de preços no mercado de energia brasileiro, saindo de uma formação explícita por meio de modelos para um novo paradigma que leve mais em consideração o poder das forças de mercado.

Continuando nesse assunto, (BASTOS et. al., 2019) propuseram uma metodologia para a separação da comercialização de energia e lastro no Brasil através da captura do valor econômico da escassez no mercado de eletricidade, buscando determinar um modelo de contratação de lastro no setor elétrico. Tal demanda se deve ao fato de que, no mercado brasileiro, determina-se a comercialização conjunta de dois produtos distintos: energia e lastro. Enquanto o primeiro é a energia elétrica em si, a ser gerada e consumida, o segundo se refere a um valor adicional a ser recebido por geradores pela capacidade de geração de energia em momentos de necessidade, o que também é comumente chamado de produto de confiabilidade. Dessa forma, faz-se necessário permitir que os agentes incorporem às suas ofertas às suas expectativas de remuneração no mercado de energia.

Em (CAMARGO et. al., 2019) foi proposta uma estratégia ótima de comercialização para consumidores livres considerando o *trade-off* entre contratação imediata e postergação de decisão, com base nas incertezas nos preços de curto prazo e na precificação de contratos bilaterais, resultantes de condições estruturais e sistêmicas que impactam de forma direta na formação do preço. Desta forma, foi implementado um modelo de otimização estocástica para a definição da estratégia de contratação para consumidores livres, no qual o controle de risco é realizado pela técnica CVaR (*Conditional Value at Risk*).

Com relação à gestão de risco na comercialização de energia, (GOMES et. al., 2019) realizaram uma avaliação de risco de exposição ao mercado de curto prazo na ocasião da aquisição de novos ativos de energia. Desta forma, as alternativas de investimento são avaliadas individualmente e combinadas, de forma a proporcionar uma comparação da receita esperada e do risco associado. A utilização da metodologia proposta permite a avaliação de alternativas não só quanto aos seus custos, mas também é possível quantificar o risco associado à escolha de determinada alternativa. Os resultados, considerando simulações realizadas para o portfólio de um gerador composto por usinas hidráulicas, contratos de venda e usinas eólicas mostraram, por exemplo, que seria desejável aumentar os recursos desse portfólio dado que os cenários analisados apresentaram preços *spot* crescentes.

Por sua vez, o trabalho apresentado em (AMADEU, 2020) objetivou a aplicação da Simulação de Monte-Carlo aliada ao Movimento Browniano Geométrico e a decomposição de Cholesky, procedimentos estocásticos amplamente utilizados nos mercados financeiro e de ações, para simular curvas estocásticas de preços futuros para o mercado brasileiro de energia, com foco no submercado Sudeste/Centro-Oeste, que corresponde à maior parte da comercialização de energia no país. Os resultados obtidos se mostraram robustos mesmo diante dos períodos de alta volatilidade observado nos últimos anos, abrangendo crises climática, econômica e sanitária, que impactaram sobremaneira a volatilidade dos preços.

Em (ARAÚJO, 2020) foi proposto um modelo baseado em dois agentes representativos que administram sua exposição ao risco através de um contrato por diferenças entre o preço futuro esperado da energia elétrica no submercado sudeste e um preço de referência de forma a abranger todos os participantes do mercado tais como comercializados e especuladores. O resultado alcançado pelo modelo foi satisfatório, considerando que a única informação utilizada são os preços *spot* esperados, informação amplamente disponível no mercado futuro de eletricidade.

Assim, ao tratar-se da previsão de preços no mercado de energia elétrica, usualmente dois termos são evidenciados pela literatura especializada: o *electricity power forecasting* (EPF) e a modelagem da curva *forward*. Apesar de ambos os termos guardam uma similaridade entre si, já que tratam de mecanismos de previsão de preços da energia elétrica, diferem entre si substancialmente devido a alguns pontos: i – a existência de prêmios de risco que variam com o tempo de maturidade do contrato; ii – a elevada incerteza na realização de previsões de preço *spot* para horizontes de tempo muito longos (LAMPS, 2021).

Além disso, em (MONTEIRO et. al., 2022) foi implementado um modelo semi-paramétrico visando obter os preços de contratos de duração diária, denominados contratos elementares, cujo portfólio constitui o período de entrega dos contratos observados no mercado. A metodologia proposta consiste em três etapas principais: a primeira consiste em um modelo estrutural semi-paramétrico que obtém as menores peças que compõem o período de entrega de um *swap* de eletricidade. Em seguida ocorre a redução da dimensionalidade do conjunto de resíduos via algoritmo PCA (Análise de Componentes Principais) e, posteriormente é realizada a estimação de um modelo de séries temporais nas componentes resultantes.

Em (BALAN et. al., 2022) foi proposto um sistema de decisão com o objetivo de definir a estratégia ótima de comercialização constituindo por módulos integrados de cenarização de geração e preços via clusterização, cálculo do custo marginal de expansão, precificação dos contratos, balanço estático, projeção de tarifa e otimização de portfólio de contratos. Desta forma, a estratégia de decisão ótima de comercialização se baseia em aspectos que vão desde a projeção da geração até a culminação com a definição dos contratos a serem transacionados.

Nesse mesmo sentido, (MARQUES et. al., 2022) propuseram um modelo de apoio à tomada de decisão sobre operações de hedge na comercialização de energia elétrica. Em geral, as operações de Hedge funcionam como um mecanismo contra a flutuação do preço da energia, contribuindo para a gestão de riscos e, conseqüentemente, trazendo maior previsibilidade de resultados para as companhias. Para isso, os autores propuseram um modelo baseado em uma função de preferência que permite modelar a variação do nível de aversão ao risco de um agente considerando diferentes bandas de preferência, mensurando o risco por meio das medidas VaR, CVaR ou ES (*Expected Shortfall*). Os resultados do modelo indicam, então, o preço ótimo de compra além do montante ideal de energia.

Ainda com relação à gestão de risco, em (CARVALHO et. al., 2022) foi proposta uma metodologia para precificar contratos de derivativos do tipo por quantidade e *collar* (de parâmetros piso, teto e ágil), indexados ao PLD, de forma que sejam traçadas curvas de mesma preferência para as possíveis combinações dos parâmetros, permitindo, desta forma, o estabelecimento de estratégias de negociação de comercializadores frente aos consumidores, levando em consideração o perfil de risco de ambas as partes envolvidas. Para análise dos resultados, foram obtidas três curvas de mesma preferência para as combinações desses parâmetros que possibilitassem ao agente comercializador ter uma ferramenta de apoio para o auxílio na negociação dos contratos.

A previsão de preços de energia elétrica desempenha um papel crucial na operação e no planejamento do setor energético, permitindo a otimização da alocação de recursos, a definição de estratégias comerciais e a mitigação de riscos financeiros para geradores, distribuidores e consumidores finais. A volatilidade dos preços no mercado de energia é influenciada por uma série de fatores exógenos, como variações na demanda, mudanças climáticas, disponibilidade de fontes renováveis, custos de geração e fatores regulatórios (WERON, 2014). Diante desse cenário, técnicas avançadas de previsão tornam-se essenciais para garantir a confiabilidade e a eficiência da gestão energética.

Historicamente, a previsão de séries temporais no setor elétrico tem sido amplamente baseada em modelos estatísticos devido à sua interpretabilidade e fundamentação matemática.

Esses modelos assumem que os valores futuros podem ser estimados a partir de padrões históricos, explorando dependências temporais bem definidas. Dentre os modelos estatísticos mais utilizados, destacam-se os baseados em processos estocásticos, como o Auto-Regressive Integrated Moving Average (ARIMA) e sua versão sazonal, o Seasonal ARIMA (SARIMA). O modelo ARIMA combina três componentes principais: autorregressivo (AR), que representa a relação entre valores passados da série; integração (I), que trata a não

estacionariedade dos dados; e médias móveis (MA), que modelam a dependência entre os resíduos das previsões passadas (BOX et al., 2015). Essa abordagem permite capturar tendências e padrões de curto prazo, sendo amplamente aplicada para previsão de preços no mercado de energia (HAMED et al., 1995).

Quando há padrões sazonais explícitos na série temporal, como oscilações diárias, semanais ou anuais na demanda e no preço da energia, o modelo SARIMA se torna uma alternativa mais robusta, pois adiciona termos sazonais às componentes do ARIMA, melhorando a capacidade preditiva em séries com periodicidades bem definidas (HYNDMAN; ATHANASOPOULOS, 2018). Estudos indicam que modelos ARIMA e SARIMA oferecem desempenho satisfatório para previsões de curto prazo em condições estáveis, mas apresentam limitações quando há eventos externos abruptos, mudanças estruturais ou alta volatilidade de preços (WERON, 2014).

Além dos modelos ARIMA e SARIMA, outras abordagens estatísticas têm sido aplicadas na previsão de preços de energia, incluindo modelos de suavização exponencial e técnicas baseadas em regressão múltipla (BENTSIKAS et al., 2020). No entanto, esses métodos frequentemente assumem que as relações entre variáveis são lineares e estacionárias, o que pode comprometer a acurácia preditiva quando a série temporal exibe comportamentos complexos e não lineares. Essa limitação tem motivado a adoção de métodos híbridos e técnicas de aprendizado de máquina, que permitem capturar padrões não lineares e melhorar a robustez da previsão em mercados dinâmicos (LAGO et al., 2021).

Nos últimos anos, o avanço da inteligência artificial tem impulsionado o uso de modelos baseados em redes neurais recorrentes (RNNs), como Long Short-Term Memory (LSTM) e Gated Recurrent Units (GRUs), que se destacam pela capacidade de aprender dependências temporais de longo alcance sem necessidade de pressuposições estatísticas rígidas (GOODFELLOW et al., 2016). Além disso, algoritmos baseados em árvores de decisão, como o Extreme Gradient Boosting (XGBoost), têm sido aplicados na previsão de preços de energia por sua eficiência computacional e habilidade de modelar interações complexas entre variáveis preditoras (CHEN; GUESTRIN, 2016).

Dessa forma, embora os modelos estatísticos tradicionais continuem sendo amplamente utilizados devido à sua interpretabilidade, sua limitação em lidar com dinâmicas não estacionárias e variações abruptas de mercado tem impulsionado o desenvolvimento de abordagens híbridas. A combinação de técnicas estatísticas, como SARIMA, com métodos baseados em aprendizado de máquina, como XGBoost, tem se mostrado uma solução promissora para melhorar a precisão preditiva e fornecer previsões mais confiáveis em cenários de alta volatilidade (AMJAD; SHAH, 2019).

2.7 Métodos Estatísticos Tradicionais

Os modelos estatísticos tradicionais desempenham um papel essencial na previsão de séries temporais, especialmente em domínios como o setor energético, onde a análise de padrões históricos pode fornecer insights valiosos para a tomada de decisão. Entre esses

métodos, o Auto-Regressive Integrated Moving Average (ARIMA) é um dos mais amplamente utilizados devido à sua robustez e aplicabilidade em diversas áreas (BOX et al., 2015). O modelo ARIMA combina três componentes principais:

- Auto-Regressivo (AR): Representa a relação entre os valores passados da série e o valor atual.
- Diferenciação (I): Torna a série estacionária ao remover tendências e variações sistemáticas.
- Médias Móveis (MA): Modelar a dependência entre os erros da previsão e os erros anteriores, reduzindo a autocorrelação dos resíduos.

O ARIMA é particularmente eficaz para séries temporais estacionárias, nas quais as propriedades estatísticas, como média e variância, permanecem relativamente constantes ao longo do tempo (BOX et al., 2015). No entanto, muitas séries temporais, incluindo preços de energia elétrica, apresentam padrões sazonais recorrentes que não são capturados de forma eficaz pelo ARIMA tradicional.

Para lidar com esse desafio, foi desenvolvido o Seasonal ARIMA (SARIMA), que estende o ARIMA ao incorporar componentes sazonais adicionais. O SARIMA adiciona termos específicos para modelar padrões periódicos previsíveis, como oscilações diárias, semanais e sazonais, tornando-o mais adequado para séries que exibem flutuações regulares ao longo do tempo (BILLINGS; YANG, 2006). Essa abordagem tem sido amplamente utilizada na previsão de séries temporais de consumo e preços de energia elétrica, pois consegue capturar fatores sazonais como ciclos climáticos, variações na demanda e eventos sazonais específicos (WANG et al., 2013).

Apesar de sua popularidade, os modelos ARIMA e SARIMA possuem limitações significativas. A principal restrição está na suposição de linearidade, o que os torna inadequados para capturar relações não lineares e dinâmicas que frequentemente ocorrem em mercados complexos, como o de energia elétrica (CHEN; GUESTRIN, 2016). Esse mercado é influenciado por uma série de fatores interdependentes, incluindo oferta e demanda, políticas regulatórias, preços de combustíveis e variáveis macroeconômicas. Quando essas interações não podem ser adequadamente modeladas por uma combinação linear de variáveis passadas, a precisão preditiva dos modelos estatísticos tradicionais pode ser comprometida (WERON, 2014).

Diante dessas limitações, pesquisadores e especialistas têm explorado abordagens alternativas que combinam métodos estatísticos com técnicas de aprendizado de máquina. Modelos híbridos que integram SARIMA com algoritmos como Extreme Gradient Boosting (XGBoost) e Redes Neurais Recorrentes (RNNs) vêm demonstrando melhorias significativas na previsão de séries temporais de preços de energia, permitindo capturar tanto padrões sazonais previsíveis quanto dinâmicas não lineares mais complexas (LAGO et al., 2021).

2.8 Métodos Baseados em Aprendizado de Máquina

Os avanços recentes em aprendizado de máquina têm impulsionado o desenvolvimento de modelos mais sofisticados para previsão de séries temporais. Diferentemente dos modelos estatísticos tradicionais, que assumem relações lineares nos dados, as técnicas de aprendizado de máquina conseguem capturar padrões não lineares e interdependências complexas, proporcionando previsões mais precisas e adaptáveis (HOCHREITER e SCHMIDHUBER, 1997).

Dentre as abordagens mais promissoras, destacam-se as redes neurais artificiais, como Redes Neurais Recorrentes (RNN) e Long Short-Term Memory (LSTM), além dos algoritmos baseados em árvores de decisão, como o Gradient Boosting Machines (GBM) e seu aprimoramento mais eficiente, o Extreme Gradient Boosting (XGBoost) (CHEN e GUESTRIN, 2016).

2.8.1 Algoritmos de Boosting e XGBoost

Os algoritmos baseados em boosting são um dos métodos mais eficazes para previsão de séries temporais em mercados dinâmicos. A técnica de boosting consiste em treinar múltiplos modelos de aprendizado sequencialmente, onde cada novo modelo corrige os erros cometidos pelo anterior (FRIEDMAN, 2001).

O XGBoost (Extreme Gradient Boosting) é uma das versões mais avançadas desse método. Ele melhora o desempenho dos modelos baseados em árvores de decisão ao incorporar regularização, paralelismo computacional e técnicas de otimização para minimizar o erro preditivo (CHEN e GUESTRIN, 2016). No contexto da previsão de preços de energia, estudos recentes demonstraram que o XGBoost é altamente eficaz para capturar variações não lineares nos dados e identificar padrões residuais que modelos estatísticos tradicionais não conseguem modelar adequadamente (JIANG et al., 2017).

O grande diferencial do XGBoost é sua capacidade de lidar com grandes volumes de dados e identificar relações ocultas nos padrões históricos. Isso o torna ideal para complementar modelos como o SARIMAX, onde os resíduos da modelagem estatística podem ser refinados por um modelo mais flexível baseado em aprendizado de máquina (WANG et al., 2014).

2.9 Modelos Híbridos na Previsão de Séries Temporais

A combinação de modelos estatísticos e de aprendizado de máquina tem sido uma abordagem amplamente estudada na previsão de séries temporais. Esses modelos híbridos aproveitam as vantagens de cada técnica para melhorar a precisão e a estabilidade das previsões (WANG et al., 2013).

O modelo SARIMAX-XGBoost é um exemplo de abordagem híbrida bem-sucedida. Ele combina a robustez do SARIMAX na modelagem de tendências e sazonalidades com a capacidade do XGBoost de capturar padrões residuais complexos. Estudos recentes apontam

que essa metodologia supera modelos individuais em termos de precisão e estabilidade preditiva (JIANG et al., 2017).

Pesquisas indicam que modelos híbridos são especialmente úteis em mercados voláteis, como o setor energético, onde múltiplos fatores influenciam os preços e o consumo (BILLINGS e YANG, 2006). Ao integrar diferentes abordagens, os modelos híbridos conseguem reduzir os erros preditivos e melhorar a capacidade de adaptação a cenários dinâmicos.

Tabela 1 – Tabela contribuições de trabalhos.

Autor (Ano)	Tema Principal	Metodologia / Técnica	Contribuições Principais
Pilopovic (1998)	Curvas forward	Teórica / Financeira	Introduziu variáveis financeiras para preços futuros
Bessembinder & Lemon (2002)	Prêmio de risco	Modelo de equilíbrio	Prêmio depende da variância e assimetria dos preços
Fleten & Fleming (2003)	Curvas forward	Preços de contratos a termo	Mostram necessidade de modelagem financeira mesmo com despacho central
Barroso et al. (2007)	PLD vs. expectativas de mercado	Análise crítica	Apontam necessidade de modelos mais adequados para médio e longo prazo
Rodolfo et al. (2015)	Gestão de risco	Pesquisa com agentes	Uso crescente de curvas forward
Lemea et al. (2015)	Análise de contratos no ACL	Preços de fechamento	Curvas de curto prazo têm maior variação
Yousefi et al. (2019)	Previsão de preços com ML	Machine Learning (78 variáveis)	Identificaram variáveis-chave para previsão de longo prazo
Mendonça et al. (2019)	Comparação de modelos preditivos	SARIMAX, LSTM, GRU, CNN-LSTM	GRU e SARIMAX tiveram maior precisão
Pappas & Santos (2019)	Formação de preço por oferta	Análise de poder de mercado	Contratos bilaterais influenciam formação de preços

Bastos et al. (2019)	Separação energia e lastro	Proposta metodológica	Nova forma de remuneração de lastro no mercado
Camargo et al. (2019)	Estratégia de contratação	Otimização estocástica com CVaR	Modelo para tomada de decisão frente à incerteza
Gomes et al. (2019)	Avaliação de ativos e risco de mercado	Análise de risco em aquisição de ativos	Integração custo e risco nas decisões de investimento

Fonte: Autores, 2025.

3. METODOLOGIA

Nesta seção, é apresentada a base teórica e a descrição geral do modelo híbrido SARIMAX-XGBoost utilizado para a previsão de preços de compra de energia elétrica a curto prazo. A abordagem proposta combina a robustez de modelos estatísticos com a flexibilidade do aprendizado de máquina, permitindo capturar tanto componentes sazonais e tendências quanto padrões residuais complexos.

O modelo proposto é estruturado em duas etapas principais: a decomposição da série temporal por meio do modelo *Seasonal Autoregressive Integrated Moving Average with exogenous variables* (SARIMAX) e a modelagem dos resíduos não explicados utilizando o algoritmo *Extreme Gradient Boosting* (XGBoost). Essa integração permite uma análise mais granular e eficiente das séries temporais de preços de energia.

3.1 Descrição do Modelo Híbrido SARIMAX-XGBoost

A implementação do modelo inicia-se com o pré-processamento dos dados, conforme ilustrado na **Figura 14**. Nessa etapa, os preços semanais de energia elétrica são normalizados pelo **MinMaxScaler**, projetando-os em um intervalo uniforme. Essa padronização evita que diferenças de magnitude entre variáveis distorçam o ajuste dos modelos e melhorem a estabilidade numérica nas etapas seguintes.

Em seguida, a série normalizada é decomposta em três componentes aditivos por meio da técnica **STL (Seasonal–Trend decomposition using Loess)**:

- **Tendência:** captura mudanças graduais e de longo prazo no comportamento dos preços, associadas a fatores macroeconômicos e regulatórios.
- **Sazonalidade:** identifica ciclos anuais recorrentes, como variações climáticas ou padrões de demanda, com periodicidade de 52 semanas.
- **Resíduo:** reúne flutuações de curto prazo e dinâmicas não lineares não explicadas pelos componentes estruturais, incluindo ruídos de mercado e eventos exógenos.

Essa decomposição organiza os dados de forma mais compreensível e viabiliza um tratamento direcionado de cada componente, elevando a precisão preditiva do modelo.

Essa decomposição organiza os dados de forma mais compreensível e viabiliza um tratamento direcionado de cada componente, elevando a precisão preditiva do modelo.

Os componentes de **Tendência** e **Sazonalidade** são então modelados pelo **SARIMAX**, cuja formulação geral, incorporando termos não-sazonais (p, d, q) , sazonais (P, D, Q) e período s , é dada por:

$$\Phi p(L)\Phi P(Ls)(1 - L)d(1 - Ls)Dyt = c + \Theta q(L)\Theta Q(Ls)\epsilon t,$$

onde

- L é o operador defasagem,
- $s = 52$ semanas,
- ϵt é ruído branco,
- Φ e Θ são polinômios autoregressivos e de média móvel (não -sazonal e sazonal),
- c é constante de intercepto.

Por fim, o componente **Resíduo** (R_t) é modelado via **XGBoost**, que explora lags de R_t , medidas de volatilidade móvel e variáveis dummy de mês para capturar interações não lineares e características latentes. A previsão desse componente é $R_t = XGB.predict(featurest)$.

Na etapa de recomposição, somam-se as previsões de cada parte para obter a série completa prevista:

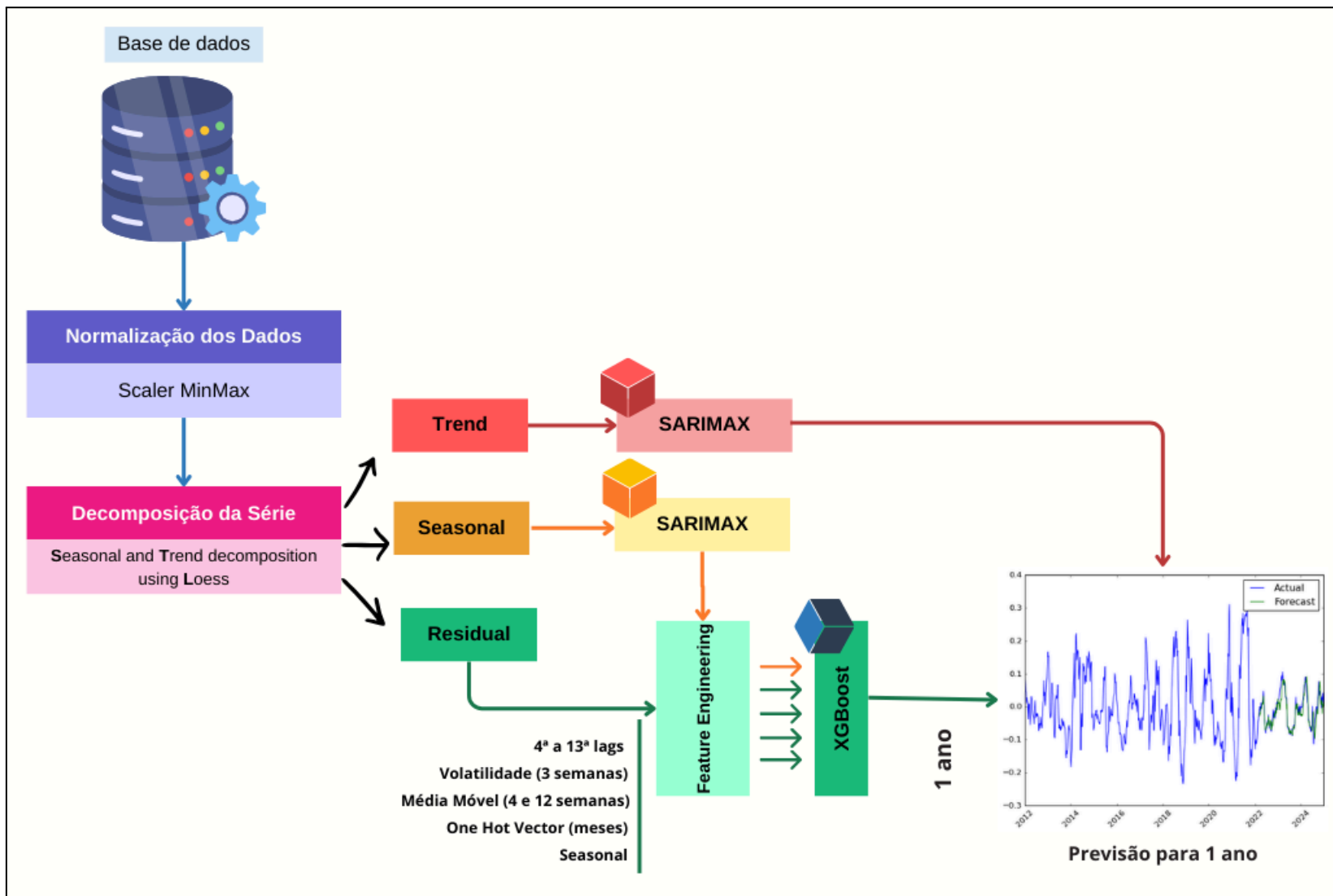
$$y_t = T_t + S_t + R_t.$$

Esse framework híbrido combina a robustez do SARIMAX na modelagem de padrões estruturais de longo prazo com a flexibilidade do XGBoost na captura de flutuações residuais, resultando em projeções mais precisas e com menor viés para horizontes de curto prazo.

3.2 Fluxo de Implementação

O modelo híbrido SARIMAX-XGBoost foi desenvolvido para abordar as complexidades e desafios na previsão de preços de compra de energia elétrica, um mercado caracterizado por alta volatilidade e padrões dinâmicos. O fluxo metodológico, ilustrado na **Figura 13**, apresenta cada uma das etapas descritas, desde o pré-processamento e a decomposição da série até a recomposição final. Os resultados demonstram que essa integração apresenta um bom desempenho para prever preços de energia elétrica, atendendo às demandas do mercado por previsões confiáveis e robustas.

Figura 13 - Fluxo de implementação do modelo híbrido SARIMAX-XGBoost para previsão de preços de energia.



Fonte: Autores, 2024

3.3 Conjunto de Dados

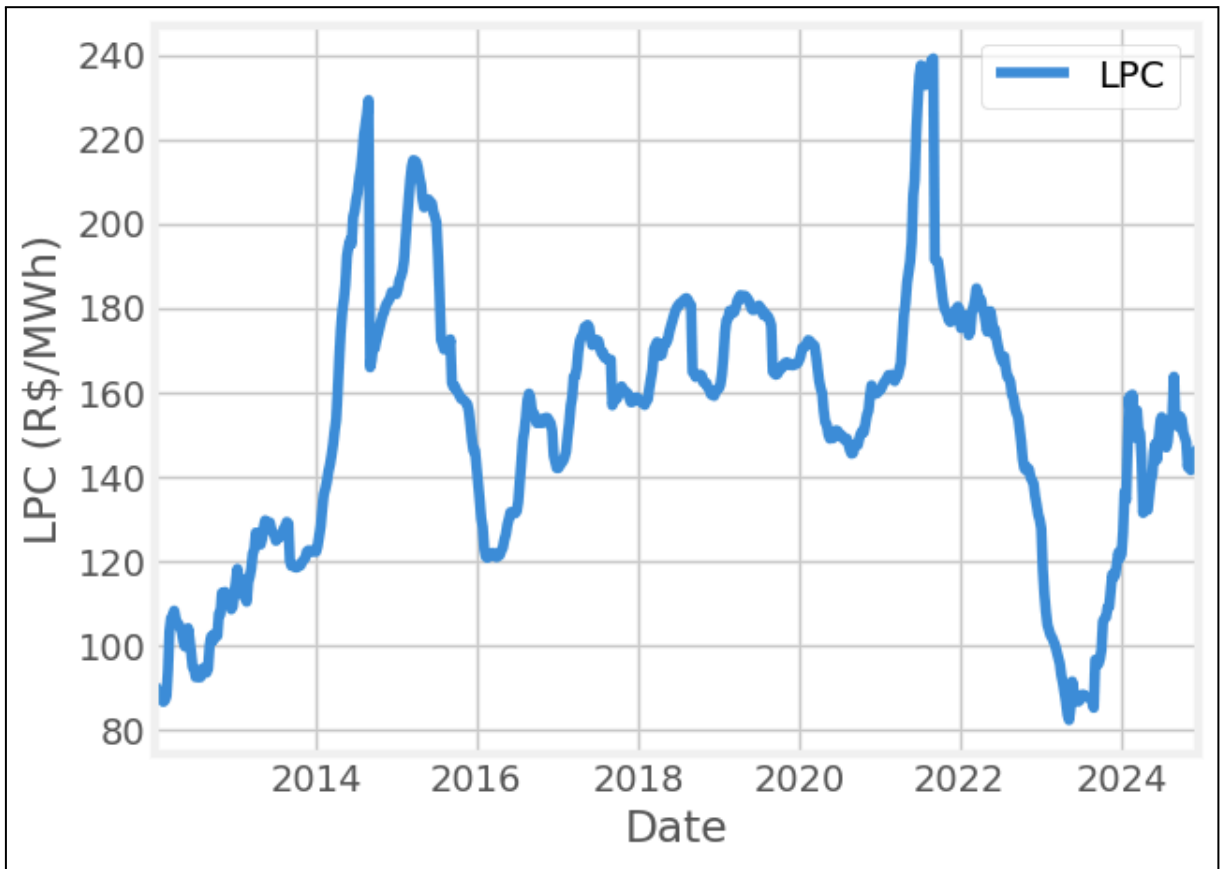
Para avaliar a eficiência e a precisão do modelo híbrido SARIMAX-XGBoost, foram utilizados dados históricos de preços de compra de energia elétrica no mercado brasileiro. O conjunto de dados abrange um período de 12 anos, de 2012 a 2024, com frequência semanal, totalizando 677 registros. Esses dados referem-se ao índice de preços de compra de energia elétrica Convencional Longo Prazo (LPC), amplamente utilizado no setor energético como referência para contratos de fornecimento em horizontes extensos. O índice LPC reflete os preços médios de energia elétrica convencionais, calculados considerando as flutuações sazonais e as tendências de longo prazo observadas no mercado.

Os dados utilizados neste estudo foram obtidos a partir dos índices apresentados no boletim de preços da Dcide, uma plataforma especializada em análises e informações do mercado de energia elétrica. A Dcide calcula semanalmente as métricas de preços do Pool com base na curva Forward, um instrumento que estima os preços futuros da energia elétrica a partir das expectativas dos agentes mais ativos comercialmente. A curva Forward captura as projeções de mercado, incluindo variações sazonais, alterações na oferta e demanda e fatores econômicos externos. Esse boletim é amplamente reconhecido no setor por sua robustez metodológica e confiabilidade, sendo uma referência essencial para análises e previsões (Dcide, 2025).

A metodologia empregada na construção desses índices contempla variações em múltiplos horizontes temporais, como alterações semanais, mensais e anuais. Além disso, abrange diferentes categorias de energia, incluindo fontes convencionais e incentivadas. Especificamente, os dados utilizados neste trabalho refletem os preços de referência em R\$/MWh para períodos de curto prazo, como trimestres móveis.

A **Figura 14** apresenta a série temporal do índice LPC utilizada neste experimento, evidenciando a evolução histórica dos preços de compra de energia elétrica ao longo do período analisado. Essa visualização destaca os padrões sazonais recorrentes, tendências de longo prazo e períodos de maior volatilidade. Esses elementos foram cruciais para justificar a escolha do modelo híbrido SARIMAX-XGBoost, que integra a modelagem de componentes estruturais, como sazonalidade e tendência, com a análise de resíduos e dinâmicas não lineares. A combinação dessas abordagens possibilitou uma previsão robusta e confiável, adequada às especificidades e complexidades do mercado energético brasileiro.

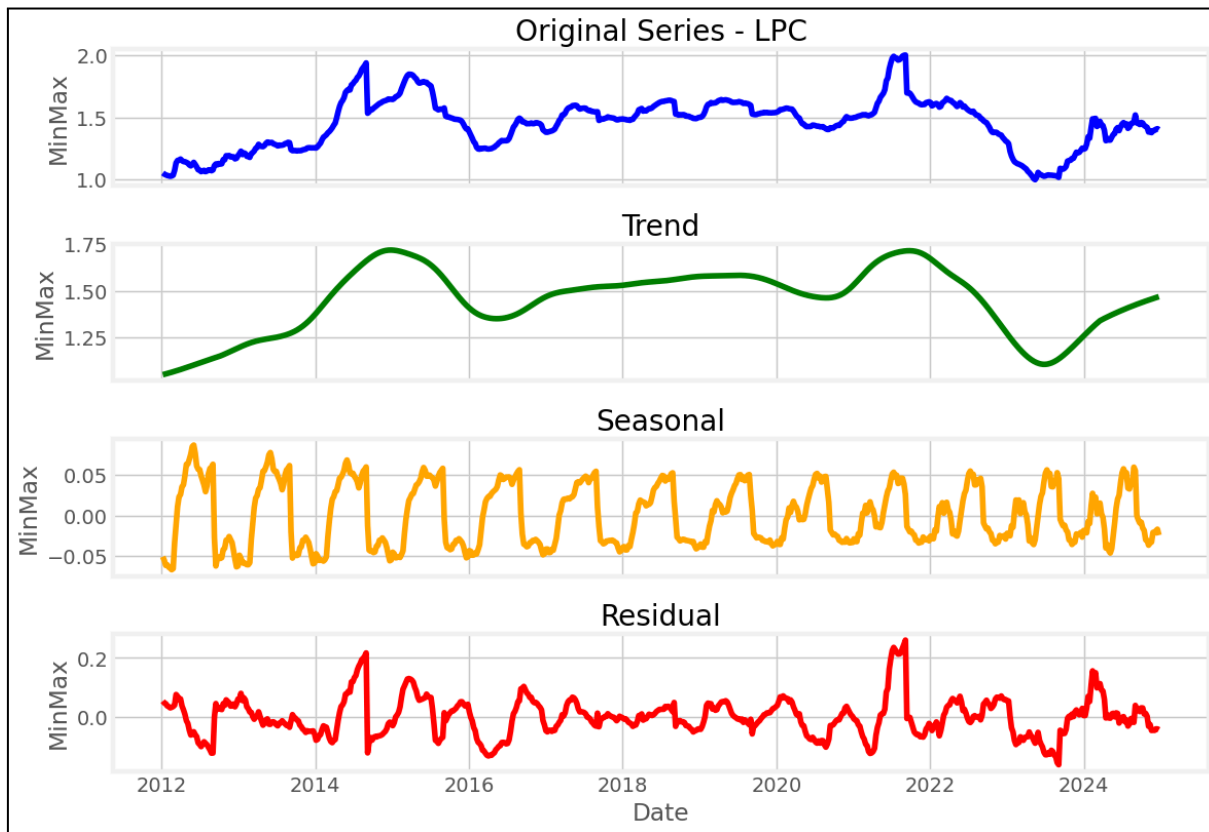
Figura 14 - Série histórica de preços Convencional Longo Prazo (LPC) de 2012 a 2024.



Fonte: Autores, 2025.

Os dados incluem variáveis como preço de compra, consumo histórico, sazonalidade e outras variáveis exógenas relevantes que influenciam diretamente o comportamento dos preços. Para analisar e modelar a série, foi realizada uma decomposição em componentes individuais, conforme ilustrado na **Figura 15**. Essa decomposição divide a série original em tendência, sazonalidade e resíduo, facilitando a modelagem e compreensão dos padrões subjacentes.

Figura 15 - Decomposição da série temporal em componentes: série original, tendência, sazonalidade e resíduo.



Fonte: Autores, 2025.

Os dados utilizados no experimento foram previamente tratados para garantir sua qualidade e adequação ao modelo proposto. O tratamento incluiu a identificação e remoção de valores anômalos por meio de regras baseadas em limites estatísticos, como o uso de z-score, permitindo a exclusão de pontos discrepantes que poderiam distorcer os resultados. Para lidar com valores ausentes, foi aplicada a interpolação linear, considerando os valores temporais adjacentes, o que garantiu a continuidade temporal dos dados sem a introdução de viés significativo. Além disso, todas as variáveis contínuas foram normalizadas, padronizando suas escalas e melhorando o desempenho dos modelos durante o treinamento.

O conjunto de dados, composto por 12 anos de informações históricas de preços de compra de energia elétrica com frequência semanal (de 2012 a 2023), foi dividido em três partes para assegurar uma avaliação robusta e imparcial do modelo híbrido SARIMAX-XGBoost. O conjunto de treinamento, representando 80% dos dados (2012 a 2020), foi utilizado para ajustar os parâmetros dos modelos e aprender os padrões estruturais e dinâmicos presentes na série temporal. O conjunto de validação, correspondente a 10% dos dados (2021), foi utilizado para a seleção de hiperparâmetros e ajustes finos, garantindo que o modelo fosse capaz de generalizar adequadamente. Por fim, o conjunto de teste, composto pelos 10% restantes dos dados (2022 a 2024), foi utilizado exclusivamente para avaliar a performance final do modelo em dados inéditos, assegurando uma análise confiável de sua precisão preditiva.

3.3.2 Pré-processamento

Os dados históricos utilizados na previsão de preços de energia elétrica podem conter registros incompletos ou valores anômalos que impactam negativamente o desempenho dos modelos. Para mitigar esses efeitos, foram aplicadas as seguintes técnicas:

- Identificação e remoção de outliers: Utilizou-se o método do Z-score para detectar valores extremos que ultrapassam três desvios-padrão da média, removendo esses registros para evitar distorções nos padrões temporais.
- Imputação de valores ausentes: Dados ausentes foram preenchidos com interpolação linear baseada nos valores adjacentes da série temporal, garantindo continuidade e evitando perdas excessivas de informações.
- Filtragem de dados inconsistentes: Foram removidos registros com variações abruptas e incoerentes, especialmente em períodos de baixa liquidez do mercado de energia.

3.3.2.1. Normalização das Variáveis

Para garantir que as variáveis tivessem escalas compatíveis e facilitar o aprendizado dos modelos, foi aplicada a normalização dos dados. A técnica utilizada foi a normalização Min-Max, definida pela equação:

$$X_{changed} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

onde $X_{changed}$ representa o valor original, X_{max} e X_{min} são os valores mínimo e máximo da variável, respectivamente. Esse procedimento foi essencial para estabilizar os valores das séries temporais e permitir uma melhor convergência dos modelos baseados em aprendizado de máquina.

3.3.2.2. Extração de Características Temporais

Para melhorar a capacidade preditiva dos modelos, foram extraídas e incorporadas ao conjunto de dados diversas características derivadas da própria série temporal, como:

- Componentes sazonais: A série foi decomposta em tendência, sazonalidade e resíduos utilizando o método STL (Seasonal-Trend Decomposition using LOESS), permitindo que o modelo lidasse separadamente com padrões de longo e curto prazo.
- Lags temporais: Foram criadas variáveis defasadas (lags) para capturar dependências temporais e permitir que os modelos aprendessem padrões recorrentes ao longo do tempo. Os lags foram configurados para intervalos de 3, 6, 9 e 12 períodos.
- Médias móveis e volatilidade: Foram adicionadas métricas como médias móveis de 5 e 12 períodos e a volatilidade do resíduo, permitindo que o modelo capturasse tendências e variações ao longo do tempo.

3.3.2.3. Codificação de Variáveis Categóricas

Como parte da análise de influência de padrões sazonais, foi realizada a transformação da variável mês do ano em formato one-hot encoding, criando colunas binárias para representar os meses de forma explícita no modelo. Essa abordagem permite que os modelos capturem efeitos sazonais mais facilmente, sem impor relações ordinais inadequadas.

3.4 Avaliação de Desempenho

A avaliação do desempenho do modelo híbrido SARIMAX-XGBoost foi conduzida utilizando métricas amplamente reconhecidas e consolidadas na análise de séries temporais. Essas métricas são ferramentas fundamentais na literatura para a análise de desempenho de previsões.

3.4.1 Métricas de Avaliação (MAE, RMSE, MAPE, R²)

Entre as métricas empregadas, destaca-se o Erro Médio Absoluto (MAE), que calcula a média dos desvios absolutos entre os valores reais e os previstos. Essa métrica é útil para quantificar a precisão geral das previsões sem penalizar desvios maiores de forma desproporcional (Billings e Yang, 2006). A Equação 1 apresenta o Cálculo do MAE.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i|$$

Onde:

n é o número de amostras

y é o valor observado para cada amostra

p é o valor previsto pelo modelo para cada amostra

|| representa o valor absoluto

O Erro Quadrático Médio (RMSE), por sua vez, dá maior peso a erros maiores, sendo amplamente utilizado para identificar desvios significativos entre os valores reais e previstos. Essa característica torna o RMSE uma métrica essencial em aplicações onde grandes discrepâncias podem comprometer a análise preditiva (BOX et al., 2015). A Equação 2 apresenta o Cálculo do RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}$$

Onde:

n é o número de amostras

y_i é o valor observado para a amostra i

p_i é o valor previsto pelo modelo para a amostra i

O Erro Percentual Absoluto Médio (MAPE) mede o erro médio relativo em termos percentuais, oferecendo uma interpretação intuitiva da precisão do modelo em relação aos valores reais. Essa métrica é particularmente útil para comparar o desempenho do modelo em diferentes escalas ou unidades de medida (WANG et al., 2014). A Equação 3 apresenta o Cálculo do MAPE.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - p_i}{y_i} \right|$$

Onde:

n é o número total de amostras

y_i é o valor real para a amostra i

p_i é o valor previsto para amostra i

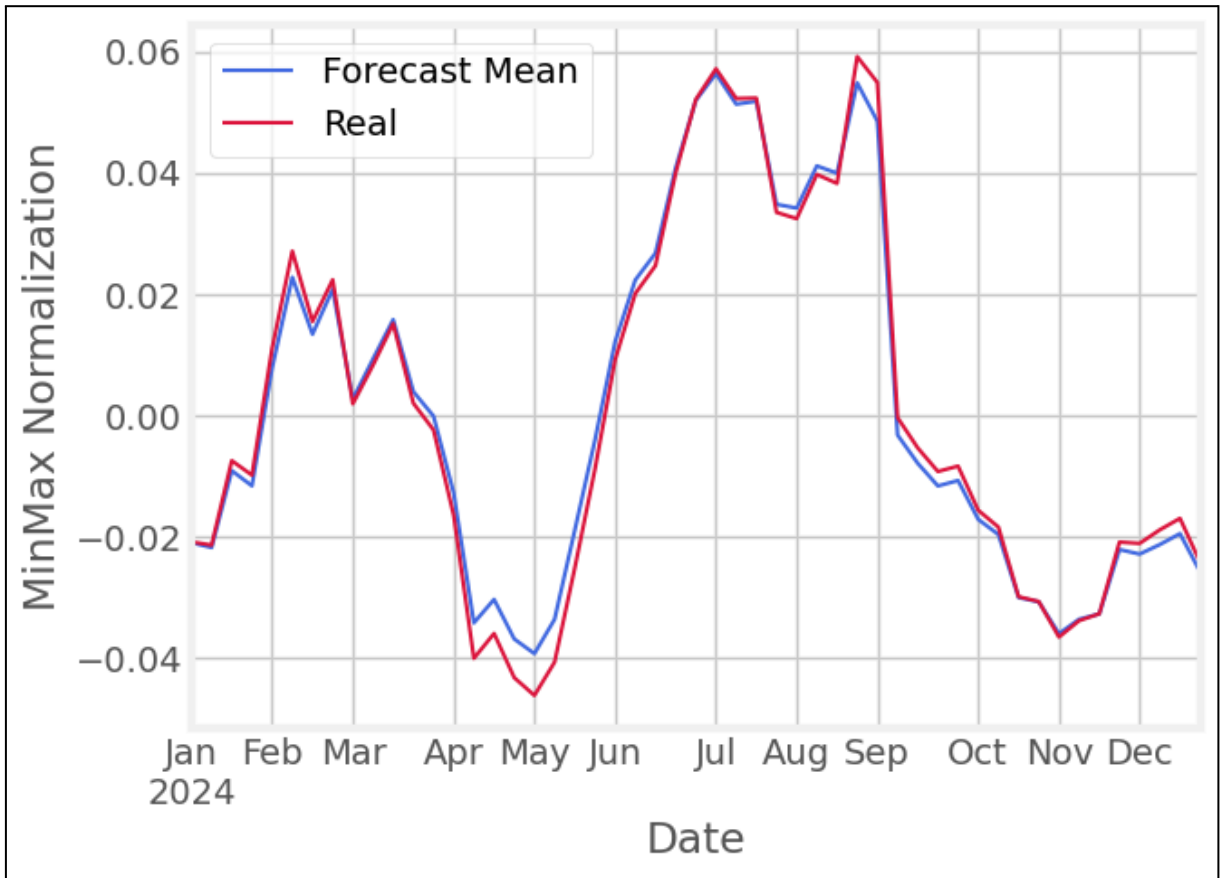
Essas métricas de avaliação garantem uma visão detalhada sobre a eficácia e as limitações do modelo híbrido SARIMAX-XGBoost na previsão de preços de compra de energia elétrica.

4. RESULTADOS E DISCUSSÃO

As previsões das componentes de sazonalidade e tendência obtidas pelo modelo SARIMAX para 2024 foram fundamentais para evidenciar padrões regulares e desvendar a dinâmica estrutural dos preços de compra de energia elétrica.

A Figura 16 destaca os ciclos sazonais semanais, com picos e vales que coincidem com variações climáticas e padrões de consumo energético típicos do setor.

Figura 16 - Projeção da Sazonalidade para o ano de 2024.

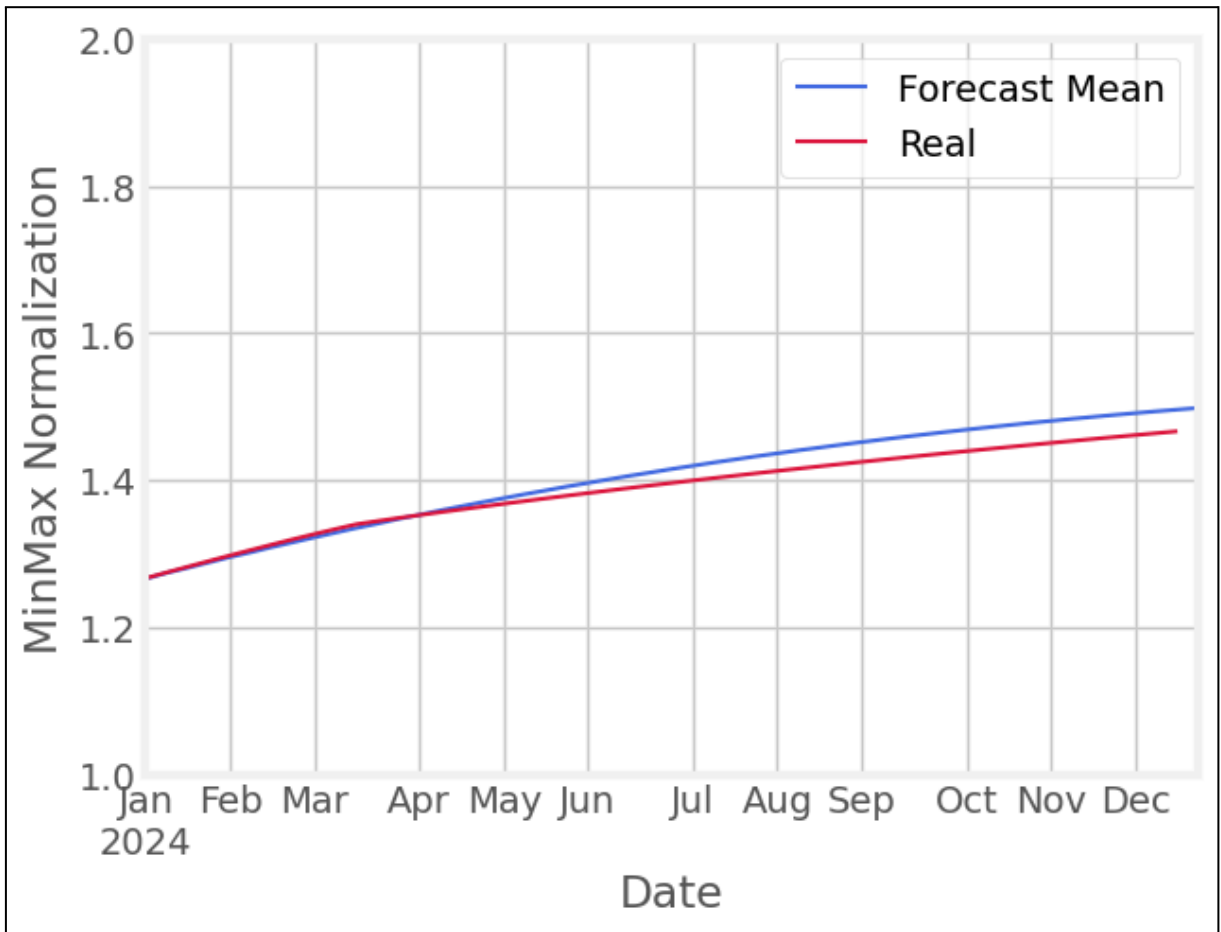


Fonte: Autores, 2025.

A previsão da série temporal para a sazonalidade e tendência pelo modelo SARIMAX projetados para o ano de 2024 foram fundamentais para identificar padrões regulares e compreender as dinâmicas estruturais subjacentes nos dados de preços de compra de energia elétrica. A projeção da sazonalidade, ilustrada na **Figura 16**, destaca flutuações regulares que ocorrem em intervalos específicos de tempo, refletindo fenômenos como variações climáticas, ciclos de consumo energético e características próprias do mercado sazonal de energia. A sazonalidade é um componente crítico, especialmente em mercados onde fatores climáticos, como estações do ano, têm um impacto direto na demanda e no preço da energia.

A projeção da tendência, apresentada na **Figura 17**, evidencia a estabilidade no preço de compra de energia elétrica. Este componente reflete mudanças graduais associadas a fatores estruturais, como crescimento da demanda, alterações no mix de geração de energia e fatores econômicos externos. A identificação clara desse comportamento é essencial para modelar a variabilidade estrutural da série temporal.

Figura 17 - Projeção da Tendência para o ano de 2024.



Fonte: Autores, 2025.

As métricas de erro obtidas para os componentes de sazonalidade e tendência na **Tabela 2** evidenciam a alta precisão do modelo SARIMAX:

Tabela 2 – Resultados da validação da projeção usando SARIMAX para Tendência e Sazonalidade.

	MAE	RMSE	MAPE
Tendência	0.0167	0.0202	1.1779
Sazonalidade	0.0023	0.0030	1.4849

Fonte: Autores, 2025.

Os resultados apresentados na **Tabela 2** mostram que o modelo SARIMAX foi altamente eficaz na modelagem dos padrões estruturais. O baixo valor do MAE indica que os desvios médios absolutos entre os valores reais e previstos são pequenos, enquanto o RMSE evidencia a capacidade do modelo de minimizar discrepâncias maiores, o que é especialmente relevante para prever oscilações significativas nos preços. O MAPE, por sua vez, fornece uma avaliação relativa, demonstrando que os erros são baixos tanto em termos absolutos quanto percentuais.

Esses resultados reforçam a capacidade do SARIMAX de capturar padrões estruturais complexos, como sazonalidade e tendências em curto prazo, ou seja, em um período de 12 meses.. Além disso, a decomposição detalhada fornece uma base sólida para a modelagem de resíduos, que são posteriormente processados pelo modelo XGBoost. A alta precisão observada nas métricas de erro indica que o SARIMAX foi eficaz em reduzir significativamente a variância não explicada na série temporal, deixando os resíduos com características mais adequadas para a próxima etapa da análise.

Os componentes de tendência e sazonalidade não apenas contribuíram para a compreensão dos padrões subjacentes na série temporal, mas também facilitaram a separação de variabilidade estrutural e residual, um passo crítico para a eficácia da abordagem híbrida. A análise visual das **Figuras 16 e 17**, em conjunto com as métricas quantitativas, oferece uma base robusta para justificar a escolha do modelo SARIMAX na etapa inicial do processo de previsão.

4.0.1 Análise dos Componentes Estruturais

Antes de recompor a série completa para previsão, efetuou-se uma decomposição estrutural via **STL (Seasonal–Trend decomposition using Loess)**, separando a série normalizada em três componentes aditivos: tendência, sazonalidade e resíduo . A **Figura 16** ilustra esse resultado, e a **Tabela 3** apresenta estatísticas descritivas de cada componente, considerando o período de 2012 a 2023.

Tabela 3 – Estatísticas descritivas dos componentes estruturais.

Componente	Média	Desvio-padrão	Amplitude típica (mín→máx)	% Variância total
Tendência	1.15	0.05	1.05 - 1.20	68%
Sazonalidade	0.00	0.04	-0.05 - +0.05	22%
Resíduo	0.00	0.015	-0.04 - +0.06	10%

Fonte: Autores, 2025.

A **Tabela 3** mostra que o componente de tendência domina a dinâmica da série, com média em 1,15 e desvio-padrão de 0,05, variando entre 1,05 e 1,20 e respondendo por 68 % da variância total. Isso indica que o comportamento de longo prazo é o principal fator de variação nos preços normalizados. Já a sazonalidade apresenta média praticamente nula, desvio-padrão de 0,04 e amplitude estreita (–0,05 a +0,05), contribuindo com 22 % da variância, o que reflete ciclos anuais regulares, porém menos expressivos que a tendência. Por fim, o resíduo tem média zero, desvio-padrão de apenas 0,015 e amplitude de –0,04 a +0,06, representando 10 % da variância total; sua baixa magnitude confirma tratar-se de flutuações de curto prazo, comparáveis a ruído branco, que exigem modelos flexíveis para captura de eventuais padrões não-lineares. Essa distribuição da variância entre os componentes reforça a escolha de um modelo SARIMAX para tendência e sazonalidade e de técnicas de machine learning para o resíduo.

4.1 Desempenho do Modelo XGBoost

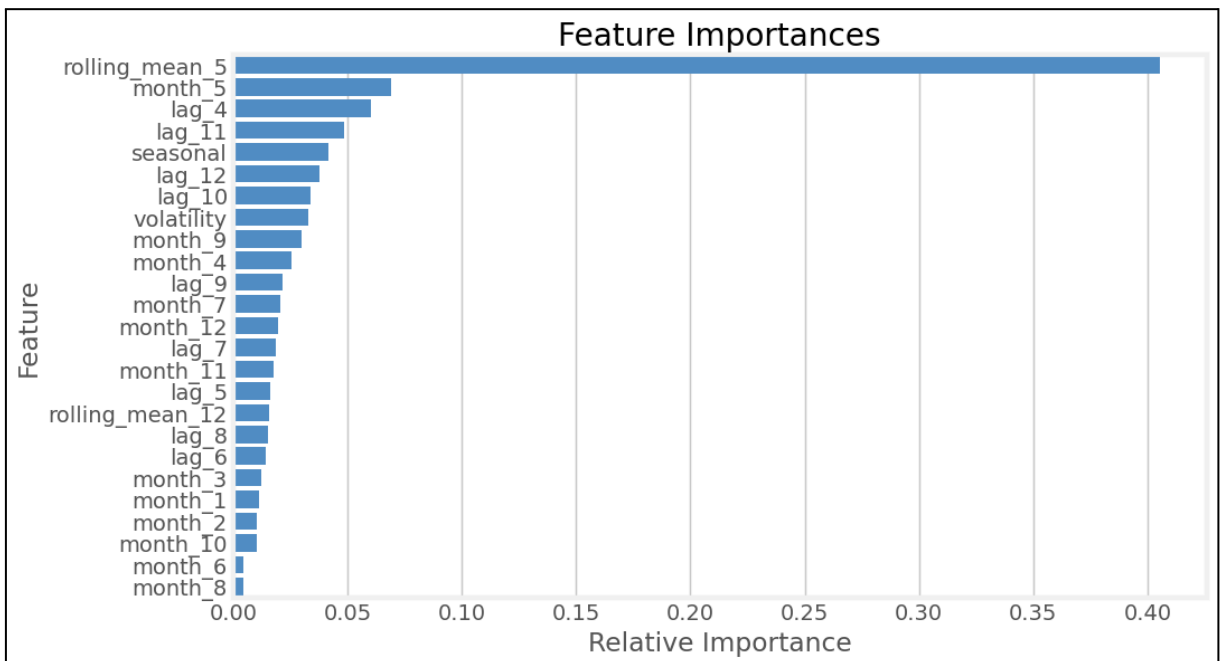
4.1.1 Análise da Importância das Variáveis

Na análise da importância das variáveis do modelo XGBoost **Figura 18**, observa-se que a média móvel de 5 períodos (`rolling_mean_5`) lidera com folga, respondendo por cerca de 40 % do ganho total de informação. Isso indica que o comportamento suavizado mais recente da série residual é o preditor mais relevante para capturar desvios não explicados pelo SARIMAX. Em seguida, a dummy referente ao mês de maio (`month_5`) e as defasagens intermediárias (`lag_4` e `lag_11`) têm grande peso, evidenciando a influência de padrões cíclicos sazonais e dependências temporais médias no ajuste final.

A presença significativa da variável “seasonal” e da defasagem de 12 períodos (`lag_12`) reforça que, mesmo depois da decomposição estatística, vestígios do ciclo anual persistem nos resíduos e demandam correções adicionais. Variáveis relacionadas à volatilidade e defasagens de curto prazo (`lag_10`, `lag_9` etc.) aparecem em posições intermediárias, sinalizando que flutuações rápidas e picos pontuais também são melhor modelados por algoritmos não lineares.

Em conjunto, essas importâncias mostram que o XGBoost, ao trabalhar sobre os resíduos do SARIMAX, equilibra três tipos de informação: o nível suavizado recente da série, dependências temporais médias e remanescentes sazonais

Figura 18 - Importância das variáveis de entrada no modelo XGBoost.

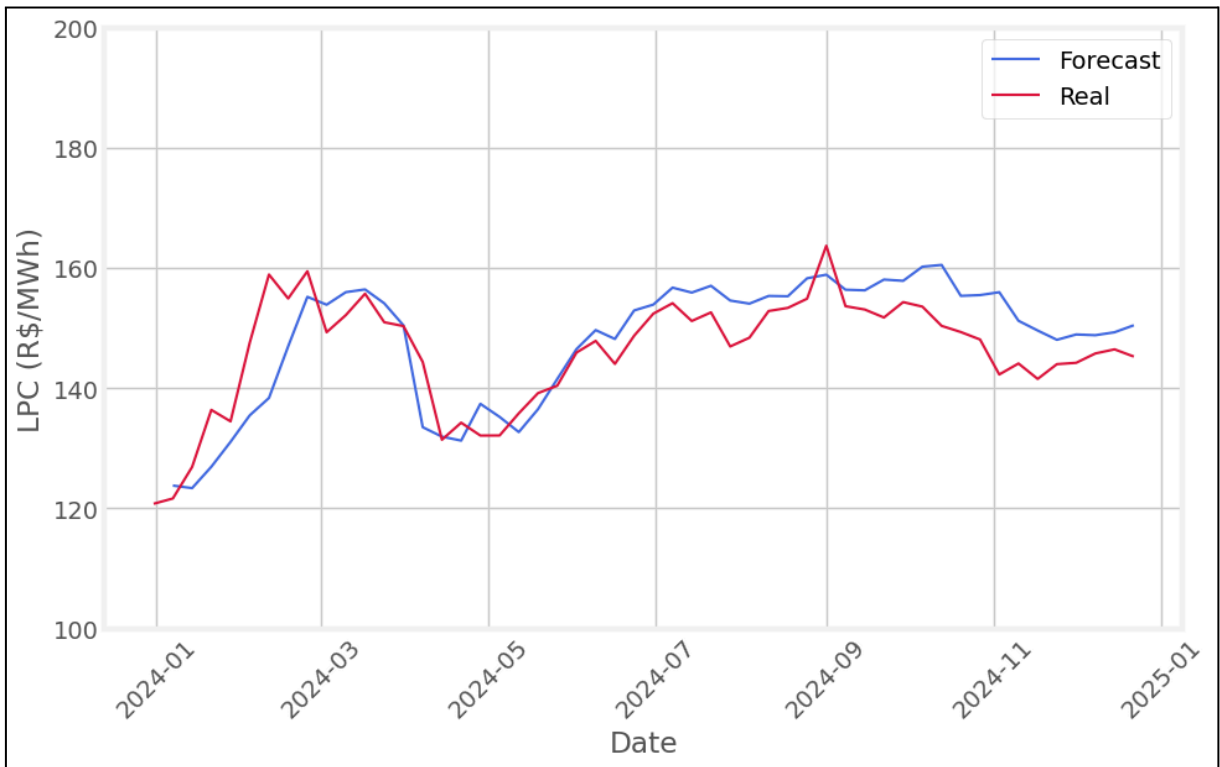


Fonte; Autores, 2025.

4.1.2 Resultados XGBoost para os resíduos.

Após a decomposição pelo modelo SARIMAX, os resíduos gerados, que representam padrões não explicados pelos componentes estruturais de tendência e sazonalidade, foram modelados utilizando o algoritmo XGBoost. Este modelo se mostrou altamente eficaz para capturar padrões residuais complexos, como dinâmicas não lineares e interações entre variáveis. A **Figura 19** apresenta a comparação entre os valores reais e previstos dos resíduos para o ano de 2024.

Figura 19 - Projeção do preço LPC para o ano de 2024.



Fonte: Autores, 2025.

Como ilustrado na **Figura 19**, o XGBoost demonstrou um desempenho consistente ao prever os resíduos, com os valores estimados mostrando uma alta correspondência com os valores observados. Essa capacidade de capturar variações residuais complexas foi crucial para complementar a análise estrutural realizada pelo SARIMAX.

As métricas de desempenho do modelo XGBoost, apresentadas na **Tabela 4**, reforçam a eficácia do algoritmo. O MAE e o RMSE evidenciam a precisão geral das previsões, enquanto o MAPE demonstra que os erros relativos médios foram baixos, indicando que o modelo é capaz de lidar com a variabilidade residual de forma robusta.

Tabela 4 – Resultados da validação da projeção usando XGBoost para o Resíduo.

	MAE	RMSE	MAPE
Resíduo	0.0212	0.0311	3.3340

Fonte: Autores, 2025.

Os resultados apresentados na **Tabela 4**, indicam que o XGBoost foi eficiente na modelagem dos resíduos, reduzindo discrepâncias não lineares e capturando relações complexas nos dados. Esse desempenho foi fundamental para recompor a série temporal com alta precisão.

4.1.3 Comparação dos métodos alternativos

A análise comparativa entre os modelos demonstra que o SARIMAX-XGBoost obteve o melhor desempenho em termos de precisão na previsão dos resíduos. A **Tabela 5** apresenta os valores das principais métricas de erro, incluindo erro absoluto médio (MAE), raiz do erro quadrático médio (RMSE) e erro percentual absoluto médio (MAPE).

Tabela 5 – Resultados da validação da projeção usando XGBoost para o Resíduo.

Model	MAE	RMSE	MAPE (%)
XGBoost	0.0212	0.0311	3.33 %
SARIMAX	0.0469	0.0604	4.29 %
LSTM	0.0336	0.0438	4.53 %
SVM	0.0150	0.0222	4.57 %

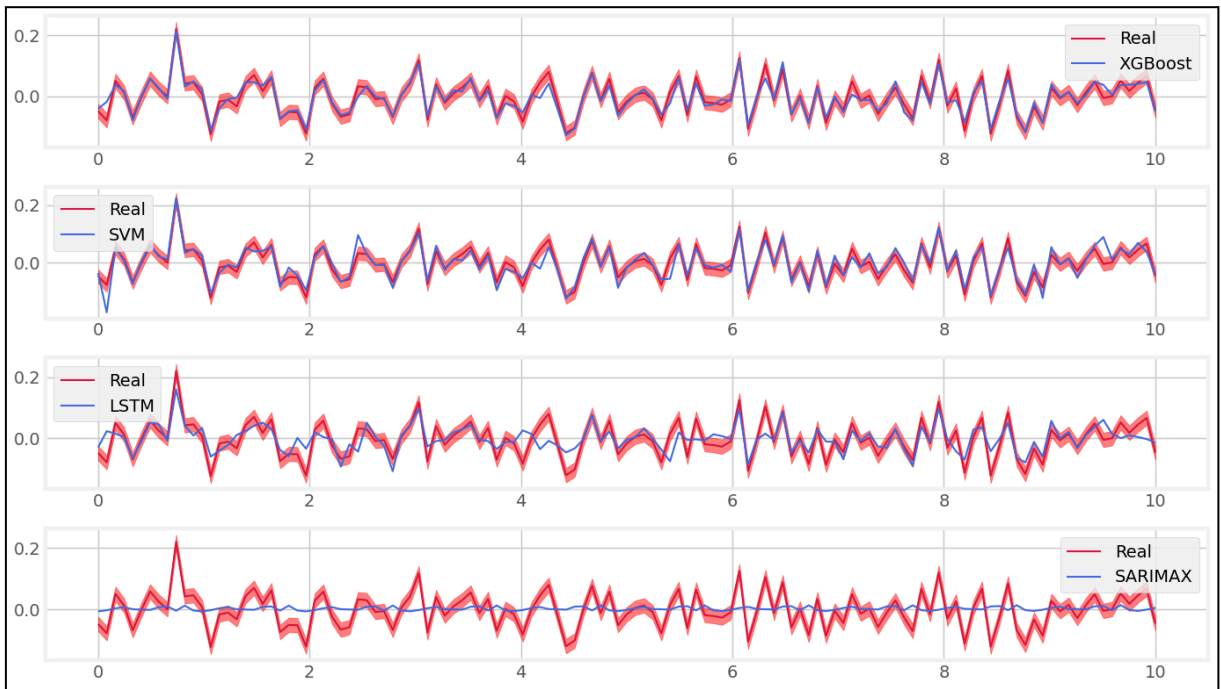
Fonte: Autores, 2025.

Os resultados demonstram que o modelo híbrido SARIMAX-XGBoost superou consistentemente os outros modelos em todas as métricas avaliadas. O modelo ARIMA apresentou o maior erro em termos de MAE, RMSE e MAPE, evidenciando sua limitação para lidar com padrões não lineares. Já o modelo LSTM teve um desempenho intermediário, reduzindo significativamente os erros em relação ao ARIMA, mas ainda inferior ao híbrido.

O modelo XGBoost puro apresentou resultados competitivos, principalmente no MAPE, refletindo sua capacidade de capturar relações não lineares nos dados. Entretanto, ao combinar a modelagem estrutural do SARIMAX com a flexibilidade do XGBoost, o modelo híbrido atingiu os melhores resultados, reduzindo o MAE em até 31,5% em comparação ao ARIMA.

A estabilidade do modelo híbrido também foi analisada por meio da distribuição dos erros absolutos, conforme ilustrado na **Figura 20**.

Figura 20 - Projeção do Resíduo para o ano de 2024.



Fonte: Autores, 2025.

A análise da Figura revela que o modelo SARIMAX-XGBoost apresentou maior estabilidade preditiva. Em contraste, os modelos ARIMA e LSTM exibiram distribuições mais dispersas, indicando maior incerteza em suas previsões.

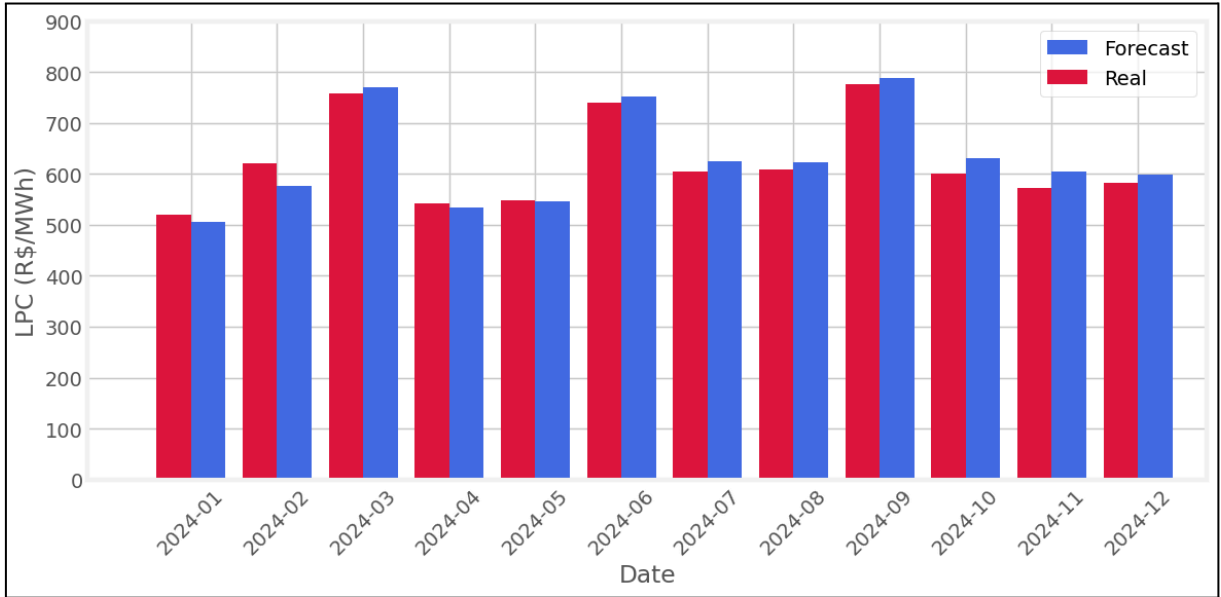
Esses resultados destacam a robustez do modelo híbrido para lidar com séries temporais complexas. Sua capacidade de capturar tanto padrões estruturais quanto dinâmicas residuais é essencial em cenários reais, como o mercado de energia, onde previsões consistentes e precisas são críticas para decisões estratégicas.

4.2 Série Recompоста e Resultados Finais

A recomposição da série temporal foi realizada pela soma dos três componentes principais: tendência, sazonalidade e resíduos. Essa abordagem híbrida permitiu capturar tanto os padrões estruturais regulares, representados pela tendência e sazonalidade, quanto às dinâmicas residuais não lineares, identificadas pelo modelo XGBoost. A integração desses elementos garantiu uma representação robusta e precisa do comportamento observado na série temporal de preços de compra de energia elétrica.

A **Figura 21** apresenta a comparação entre os valores reais e previstos da série recompоста para o ano de 2024. Essa visualização destaca a capacidade do modelo híbrido SARIMAX-XGBoost de representar fielmente as variações nos dados, minimizando discrepâncias entre os valores reais e estimados.

Figura 21 - Comparação entre os valores reais e previstos para o ano de 2024, considerando a recomposição completa.



Fonte: Autores, 2025.

As métricas de desempenho da série recomposta estão detalhadas na **Tabela 6**. Os resultados indicam a alta precisão do modelo, com valores baixos de erro absoluto médio (MAE), erro quadrático médio (RMSE) e erro percentual absoluto médio (MAPE).

Tabela 6 – Resultados após a recomposição das projeções das componentes.

	MAE	RMSE	MAPE
LPC	1.8009	2.1422	2.9459

A variável LPC combina (Tendência + Sazonalidade + Resíduo) já recomposto.

Fonte: Autores, 2025.

Os resultados evidenciam que a abordagem híbrida SARIMAX-XGBoost é altamente eficaz na previsão de preços de compra de energia elétrica. A integração dos componentes estruturais e residuais permitiu capturar tanto padrões regulares quanto variações complexas, fornecendo uma ferramenta robusta e precisa para cenários reais do mercado energético.

5. CONCLUSÃO E TRABALHOS FUTUROS

Os resultados obtidos com a aplicação do modelo híbrido SARIMAX-XGBoost evidenciam sua eficácia na previsão de preços de compra de energia elétrica no contexto brasileiro, destacando o potencial dessa abordagem para enfrentar os desafios impostos pela complexidade das séries temporais do mercado energético. A metodologia de decomposição e recomposição da série temporal, realizada pela integração dos componentes de tendência, sazonalidade e modelagem de resíduos, demonstrou que a combinação de técnicas estatísticas

e aprendizado de máquina permite capturar tanto os padrões estruturais quanto as dinâmicas residuais não lineares presentes nos dados.

O modelo SARIMAX desempenhou papel fundamental na identificação e modelagem dos padrões estruturais da série temporal. A tendência capturada representou mudanças de longo prazo, refletindo fatores econômicos, políticos e regulatórios que influenciam o mercado de energia brasileiro. A sazonalidade, por sua vez, permitiu identificar variações regulares e previsíveis, como flutuações associadas a ciclos climáticos ou períodos de alta e baixa demanda, característicos do sistema predominantemente hidrelétrico nacional. A capacidade do SARIMAX de incorporar variáveis exógenas também possibilitou a inclusão de fatores externos relevantes, como indicadores hidrológicos e econômicos, enriquecendo a modelagem dos componentes estruturais.

A decomposição da série temporal em seus componentes estruturais foi crucial para reduzir a variabilidade nos resíduos, preparando-os adequadamente para a modelagem com o XGBoost. Esta etapa intermediária representa uma contribuição metodológica significativa, pois facilita a tarefa de cada algoritmo ao permitir que se concentre em aspectos específicos da série temporal, em vez de tentar modelar toda sua complexidade simultaneamente.

O XGBoost demonstrou alto desempenho na modelagem dos resíduos, capturando padrões não lineares e interações complexas que não foram explicadas pelos componentes estruturais. A análise das métricas de desempenho do modelo, como o MAE, RMSE e MAPE, indicou que o XGBoost foi eficiente em minimizar erros e melhorar a precisão preditiva. Além disso, a análise de importância das variáveis reforçou a relevância de fatores como médias móveis e defasagens temporais na previsão dos resíduos, oferecendo insights valiosos sobre a dinâmica subjacente aos dados do mercado energético brasileiro.

A recomposição da série temporal, que integrou os componentes modelados separadamente, resultou em previsões com precisão significativamente superior aos modelos individuais. A comparação entre os valores reais e previstos, ilustrada graficamente nos resultados, evidencia a robustez do modelo híbrido em reproduzir o comportamento observado na série temporal de preços de energia. Os valores das métricas de desempenho para a série recomposta, incluindo um MAPE abaixo de 3%, destacaram a consistência do modelo em capturar as variações dos preços com alta precisão. Essa performance é particularmente relevante em um setor onde decisões estratégicas dependem de previsões confiáveis e adaptáveis às rápidas mudanças do mercado.

5.1 Síntese dos Resultados

A síntese dos resultados obtidos neste estudo pode ser estruturada em três aspectos principais: desempenho preditivo, interpretabilidade do modelo e aplicabilidade prática.

Em termos de desempenho preditivo, o modelo híbrido SARIMAX-XGBoost superou consistentemente as abordagens individuais em todas as métricas avaliadas. A redução do erro médio absoluto (MAPE) para valores inferiores a 3% representa um avanço significativo em

relação aos métodos tradicionais, que frequentemente apresentam erros na faixa de 5-10% para previsões de preços de energia no mercado brasileiro. Esta melhoria na precisão foi observada tanto em períodos de relativa estabilidade quanto em momentos de maior volatilidade, demonstrando a robustez do modelo proposto.

Quanto à interpretabilidade, a abordagem híbrida oferece vantagens significativas ao permitir a análise separada dos componentes estruturais e residuais. A decomposição realizada pelo SARIMAX possibilita identificar claramente tendências de longo prazo e padrões sazonais, facilitando a compreensão dos fatores fundamentais que influenciam os preços de energia. Simultaneamente, a análise de importância das variáveis no modelo XGBoost fornece insights sobre os fatores que afetam os componentes residuais, complementando a interpretação estrutural com uma visão das interações complexas presentes nos dados.

No que concerne à aplicabilidade prática, os resultados indicam que o modelo proposto pode ser implementado como uma ferramenta efetiva para apoiar decisões estratégicas no mercado energético brasileiro. A capacidade de prever preços com maior precisão tem implicações diretas para a gestão de riscos financeiros, planejamento de compra e venda de energia, e otimização de portfólios energéticos. A metodologia desenvolvida é suficientemente flexível para ser adaptada a diferentes horizontes de previsão e para incorporar novas variáveis exógenas conforme necessário, aumentando sua utilidade prática em um mercado em constante evolução.

5.2 Limitações do Estudo

Apesar dos resultados promissores, é importante reconhecer as limitações inerentes a este estudo. Primeiramente, a abordagem proposta foi validada utilizando dados históricos específicos do mercado brasileiro, o que pode limitar sua generalização para outros contextos ou períodos com características significativamente diferentes. Eventos extremos ou mudanças estruturais no mercado, como alterações regulatórias profundas ou crises energéticas severas, podem representar desafios adicionais para o modelo.

Uma segunda limitação refere-se à disponibilidade e qualidade dos dados utilizados. Embora tenha sido empregado um conjunto abrangente de variáveis exógenas, existem fatores potencialmente relevantes que não puderam ser incorporados devido a restrições de acesso a dados ou dificuldades de quantificação. Aspectos como decisões políticas, mudanças regulatórias antecipadas ou expectativas de mercado são exemplos de variáveis difíceis de parametrizar que podem influenciar significativamente os preços de energia.

Do ponto de vista metodológico, a abordagem de decomposição e recomposição adotada, embora eficaz, introduz complexidade adicional ao processo de modelagem e pode amplificar erros se os componentes não forem adequadamente estimados. A necessidade de ajustar múltiplos modelos também aumenta o custo computacional e a complexidade de implementação em sistemas de produção.

Finalmente, é importante considerar que o horizonte de previsão explorado neste estudo foi limitado ao curto prazo. A extensão da metodologia para previsões de médio e longo prazo pode requerer adaptações significativas e possivelmente a incorporação de técnicas adicionais para lidar com a crescente incerteza associada a horizontes temporais mais distantes.

5.3 Contribuições do Trabalho

Apesar dos resultados promissores, é importante reconhecer as limitações inerentes a este estudo. Primeiramente, a abordagem proposta foi validada utilizando dados históricos específicos do mercado brasileiro, o que pode limitar sua generalização para outros contextos ou períodos com características significativamente diferentes. Eventos extremos ou mudanças estruturais no mercado, como alterações regulatórias profundas ou crises energéticas severas, podem representar desafios adicionais para o modelo.

Uma segunda limitação refere-se à disponibilidade e qualidade dos dados utilizados. Embora tenha sido empregado um conjunto abrangente de variáveis exógenas, existem fatores potencialmente relevantes que não puderam ser incorporados devido a restrições de acesso a dados ou dificuldades de quantificação. Aspectos como decisões políticas, mudanças regulatórias antecipadas ou expectativas de mercado são exemplos de variáveis difíceis de parametrizar que podem influenciar significativamente os preços de energia.

Do ponto de vista metodológico, a abordagem de decomposição e recomposição adotada, embora eficaz, introduz complexidade adicional ao processo de modelagem e pode amplificar erros se os componentes não forem adequadamente estimados. A necessidade de ajustar múltiplos modelos também aumenta o custo computacional e a complexidade de implementação em sistemas de produção.

Finalmente, é importante considerar que o horizonte de previsão explorado neste estudo foi limitado ao curto prazo. A extensão da metodologia para previsões de médio e longo prazo pode requerer adaptações significativas e possivelmente a incorporação de técnicas adicionais para lidar com a crescente incerteza associada a horizontes temporais mais distantes.

5.4 Sugestões para Trabalhos Futuros

Com base nos resultados obtidos e nas limitações identificadas, várias direções promissoras para pesquisas futuras podem ser delineadas. Uma primeira linha de investigação envolve a extensão do modelo híbrido para horizontes de previsão mais longos, explorando técnicas específicas para capturar dependências de longo prazo e tendências estruturais em escalas temporais maiores.

Uma segunda direção de pesquisa consiste na incorporação de fontes de dados alternativas e não convencionais ao modelo. A integração de dados textuais provenientes de notícias, relatórios setoriais e mídias sociais, por exemplo, poderia enriquecer a capacidade do modelo de capturar expectativas de mercado e antecipar mudanças regulatórias. Técnicas de processamento de linguagem natural (PLN) poderiam ser empregadas para extrair sentimentos e informações relevantes desses dados não estruturados.

A adaptação do modelo para previsão probabilística, em vez de pontos estimados, representa outra área promissora para a investigação. A quantificação da incerteza associada às previsões através de intervalos de confiança ou distribuições de probabilidade completas forneceria informações valiosas para a gestão de riscos e tomada de decisão sob incerteza.

Adicionalmente, a exploração de arquiteturas de aprendizado profundo, como redes neurais recorrentes (LSTM, GRU) ou transformers, para a modelagem de resíduos poderia potencialmente melhorar ainda mais a capacidade do modelo de capturar padrões complexos e não lineares. A comparação sistemática dessas abordagens com o XGBoost utilizado neste estudo fornecem insights sobre as vantagens relativas de diferentes algoritmos de aprendizado de máquina no contexto específico da previsão de preços de energia.

Por fim, a aplicação da metodologia desenvolvida a outros mercados energéticos, tanto nacionais quanto internacionais, permitiria avaliar sua generalização e identificar adaptações necessárias para diferentes contextos regulatórios e estruturas de mercado. Estudos comparativos entre diferentes países poderiam revelar padrões comuns e especificidades locais na dinâmica de formação de preços de energia elétrica, contribuindo para uma compreensão mais abrangente deste fenômeno complexo.

REFERÊNCIAS

BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C. **Time Series Analysis: Forecasting and Control**. Wiley, 2015.

CHEN, Tianqi; GUESTRIN, Carlos. **XGBoost: A scalable tree boosting system**. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, p. 785-794. DOI: 10.1145/2939672.2939785. Disponível em: <https://doi.org/10.1145/2939672.2939785>. Acesso em: 27 jan. 2025.

FAN, S.; CHEN, L. **Short-term load forecasting based on an adaptive hybrid method**. IEEE Transactions on Power Systems, v. 21, n. 1, p. 392-401, fev. 2006. DOI: 10.1109/TPWRS.2005.860944. Disponível em: <https://ieeexplore.ieee.org/document/1409425>. Acesso em: 27 jan. 2025.

FRIEDMAN, Jerome H. **Greedy function approximation: A gradient boosting machine**. Annals of Statistics, v. 29, n. 5, p. 1189-1232, out. 2001. DOI: 10.1214/aos/1013203451. Disponível em: <https://doi.org/10.1214/aos/1013203451>. Acesso em: 27 jan. 2025.

BILLINGS, D.; YANG, J.-S. **Application of the ARIMA Models to Urban Roadway Travel Time Prediction - A Case Study**. In: 2006 IEEE International Conference on Systems, Man and Cybernetics, Taipei, Taiwan, 2006. p. 2529-2534. DOI: 10.1109/ICSMC.2006.385244. Disponível em: <https://ieeexplore.ieee.org/document/4274250>. Acesso em: 27 jan. 2025.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. Neural Computation, v. 9, n. 8, p. 1735-1780, 1997. DOI:

<https://doi.org/10.1162/neco.1997.9.8.1735>. Disponível em: <https://doi.org/10.1162/neco.1997.9.8.1735>. Acesso em: 27 jan. 2025.

JIANG, Ping; WANG, Yun; WANG, Jianzhou. **Short-term wind speed forecasting using a hybrid model**. Energy, v. 119, p. 561-577, 2017. ISSN 0360-5442. DOI: 10.1016/j.energy.2016.10.040. Disponível em: <https://www.sciencedirect.com/science/article/pii/S036054421631475X>. Acesso em: 27 jan. 2025.

WANG, H.; LIU, L.; QIAN, Z. (Sean); WEI, H.; DONG, S. **Empirical Mode Decomposition–Autoregressive Integrated Moving Average: Hybrid Short-Term Traffic Speed Prediction Model**. Transportation Research Record, v. 2460, n. 1, p. 66-76, 2014. DOI: 10.3141/2460-08. Disponível em: <https://doi.org/10.3141/2460-08>. Acesso em: 27 jan. 2025.

DCIDE. **Boletim Semanal de Preços – Semana 04 de 2025**. Disponível em: <https://www.dcide.com.br/wp-content/uploads/2025/01/Boletim-S04.25.pdf>. Acesso em: 22 jan. 2025.

AMJAD, M.; SHAH, D. **Trading Bitcoin and Online Time Series Prediction**. In: Proceedings of the 34th International Conference on Machine Learning, 2019.

BENTSIKAS, P.; TEGOU, L.; PSARRAS, J. **A Review of Electricity Price Forecasting Techniques**. Renewable and Sustainable Energy Reviews, v. 134, p. 110312, 2020.

CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System**. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016.

HAMED, K. H.; RAMIREZ, J. A.; SMITH, R. A. **Short-Term Electricity Load Forecasting Using SARIMA Model**. Journal of Water Resources Planning and Management, v. 121, n. 6, p. 414-423, 1995.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. Melbourne: OTexts, 2018.

LAGO, J.; DE RIDDER, F.; DE SCHUTTER, B. **Forecasting Spot Electricity Prices: Deep Learning Approaches and Empirical Comparison of Traditional Algorithms**. Applied Energy, v. 268, p. 114943, 2021.

WERON, R. **Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future**. International Journal of Forecasting, v. 30, n. 4, p. 1030-1081, 2014.